

Das Ende von Rational Choice? Zur Leistungsfähigkeit der Rational-Choice-Theorie

Betz, Dirk

Veröffentlichungsversion / Published Version
Dissertation / phd thesis

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Betz, D. (2021). *Das Ende von Rational Choice? Zur Leistungsfähigkeit der Rational-Choice-Theorie*. (GESIS-Schriftenreihe, 27). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.73617>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

gesis

Leibniz-Institut
für Sozialwissenschaften

Schriftenreihe

Band 27

Dirk Betz

Das Ende von Rational Choice?

Zur Leistungsfähigkeit der Rational-
Choice-Theorie

mit einem Vorwort von Hartmut Esser

Das Ende von Rational Choice? Zur Leistungsfähigkeit der Rational-Choice-Theorie

GESIS-Schriftenreihe

herausgegeben von GESIS – Leibniz-Institut für Sozialwissenschaften

Band 27

Dirk Betz

Das Ende von Rational Choice? Zur Leistungsfähigkeit der Rational-Choice-Theorie

Die vorliegende Arbeit wurde am Cologne Center for Comparative Politics (CCCP), der wirtschafts- und sozialwissenschaftlichen Fakultät der Universität zu Köln als Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.) angenommen.

Referent: Prof. Dr. André Kaiser, Universität zu Köln;

Korreferent: Prof. Dr. Martin Fochmann, Universität zu Köln

Dirk Betz

Das Ende von Rational Choice?

Zur Leistungsfähigkeit der Rational-Choice-Theorie

**mit einem Vorwort von
Hartmut Esser**

Bibliographische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://www.dnb.de> abrufbar.

ISBN	978-3-86819-046-5 (print)
ISBN	978-3-86819-045-8 (eBook)
ISSN	1869-2869

Herausgeber,

Druck u. Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln, Tel.: 0221 / 476 94 - 0
publications@gesis.org
Printed in Germany

Nutzungsbedingungen: Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Inhaltsverzeichnis

Vorwort von Hartmut Esser	9
1 Einleitung	11
2 Beitrag der Dissertation zur forschungstheoretischen Diskussion	23
3 Konditionale Kooperation in experimentellen Öffentliches-Gut-Spielen mit und ohne Feedback: Erklärung oder Artefakt?	43
4 Fixation durations – Why are they so highly variable?	93
5 Der Einfluss unterschiedlicher Gerechtigkeitsprinzipien auf Mehrheits- entscheidungen in einem spieltheoretischen Social-Good-Experiment ...	119
6 Fairness ohne Eigennutz. Handlungsannahmen in der ökonomischen Theorie und Auswirkungen auf Reformoptionen im Gesundheitswesen ..	155
7 The Squeaky Wheel Gets the Grease – Reference Point Variations in Game Theoretic Experimental Dictator Games	175

Danksagung

Ich bedanke mich bei Claudia Biniossek, Jeannette Brosig-Koch, Kathrin Busch, Hartmut Esser, Silke Fengler, Detlef Fetchenhauer, Martin Fochmann, Jürgen Friedrichs, Niels Galley, Andreas Gerber-Grote, Ivo Gross, Timo Heinrich, Ursula Hoffmann, André Kaiser, Bernhard Kittel, Ulrike Lange, Markus Längen, Rosemarie Nagel, Sebastian Netscher, Marcel Oestreich, Abdolkarim Sadrieh, Joachim Weimann, Harald Wypior und allen, die mich bei der Anfertigung dieser Arbeit unterstützt haben. Diese Arbeit ist meiner Familie und meinen Freunden gewidmet.

Vorwort

Die Frage nach der Einheit der (Sozial-)Wissenschaften war immer schon umstritten, und sie ist es derzeit, wie es aussieht, mehr denn je. Die Signale sind zwar unterschiedlich, aber nicht unbedenklich, und gerade in den Sozialwissenschaften scheint es eher wieder auseinander zu gehen: In der Soziologie, die am stärksten wohl, auch in der Politikwissenschaft und sogar in der Ökonomie inzwischen. Eine doch recht lange Zeit hatte die ökonomische Perspektive die Möglichkeit aufscheinen lassen, dass durch die Übernahme ihrer Standards und besonders ihres theoretischen Instrumentariums ein einheitlicher Rahmen geschaffen werden könnte: durch die Erweiterung und Lockerung gewisser Annahmen der Rational-Choice-Theorie, zuletzt noch bei Herbert Gintis in seinem Versuch, nun auch die Kultur und die Symbole mit aufzunehmen. Das hatte immer einen starken Anteil an eher feindlicher Übernahme gehabt. Keine gute Grundlage für eine Perspektive, die nicht auf Dominanz, sondern arbeitsteilige Differenzierung und Spezialisierungsgewinne im beständigen Austausch setzt. Kristallisiert hatte sich das an zwei Entwicklungen: Die Zunahme einer ganzen Reihe von Anomalien der traditionellen Rational-Choice-Theorie, wie die Framing-Effekte und die Entdeckung von Heuristiken aller Art. und die Befunde aus spieltheoretischen Experimenten, die die aufgetretenen Schwierigkeiten der Rational-Choice-Theorie eher bestätigten, aber auch die Ratlosigkeit, wie damit umzugehen sei, eher noch vergrößerten. Die Arbeit Dirk Betz ist vor diesem Hintergrund zu sehen. Anders als sonst meist in dieser Diskussion bringt sie in einem ganz besonderem Maße über den Versuch einer theoretischen Einordnung der Rational-Choice-Theorie in einen übergreifenden Rahmen kognitiver Prozesse, die gerade auch kulturelle und symbolische Vorgänge in eine auch formal stringente Modellierung überführt, Öffnung und Entspannung. Denn man sieht dann, wo die einzelnen Ansätze, auch die Rational-Choice-Theorie, ihren angemessenen Platz haben können. Die Arbeit kümmert sich aber dann, besonders in den Einzelbeiträgen, gleich wieder um die systematische empirische Fundierung der Überlegungen, gezielt auch in Experimenten und erweiterten Replikationen der klassischen Beiträge dazu - ganz so wie das eigentlich gar nicht anders sein sollte. Es ist ein enorm wichtiger Beitrag zu einem schon drängend gewordenen Problem. Und eine gute Grundlage für den sicher auch schwierigen weiteren Weg hin zu einer integrativen und damit zwingend interdisziplinär orientierten Sozialtheorie der Sozialwissenschaften.

1 Einleitung

Thematische Einführung

Das Verhalten von Individuen im institutionellen Handlungsrahmen als Gegenstandsbereich der vergleichenden Politikwissenschaft

Die Erklärung individuellen Entscheidungsverhaltens im institutionellen Handlungsrahmen steht im Zentrum eines breiten politikwissenschaftlichen Diskurses, der diesen Gegenstandsbereich unter folgenden Fragestellungen betrachtet: Durch welche Mechanismen wird menschliches Verhalten gesteuert? Wie erlangen Institutionen im Denken und für das Handeln von (politischen) Akteuren Bedeutung¹? Welche dynamischen, interdependenten Anpassungsprozesse haben zu den empirisch beobachtbaren institutionellen Ausprägungen geführt? Ein wachsender Strang dieser Literatur macht hierzu handlungstheoretische Ansätze für die empirisch-analytische Politikwissenschaft nutzbar. Die vorliegende Dissertation stellt als Beitrag zur Erklärung individuellen Entscheidungsverhaltens die Untersuchung der Leistungsfähigkeit unterschiedlicher handlungstheoretischer Forschungstraditionen in den Mittelpunkt.

Der zunehmende Einsatz des ökonomischen Verhaltensparadigmas in vielfältigen politikwissenschaftlichen Anwendungsgebieten seit Ende der 80er Jahre (Braun, 2013) verläuft parallel zu einer wachsenden Verbreitung des ökonomischen Verhaltensparadigmas in den Sozialwissenschaften, hier in Form der Rational-Choice-Theorie (RCT) bzw. des Public-Choice-Ansatzes. Der Erforschung individuellen Entscheidungsverhaltens im Rahmen von incentivierten Experimenten seit Mitte der 90er Jahre (Kittel, Luhan, & Morton, 2012) kommt auch in der vergleichenden Politikwissenschaft immer größere Bedeutung zu. Die RCT geht sowohl in ihren engen als auch in ihren weiten Varianten von rational-kalkulierenden Individuen aus, die einem Optimierungskalkül folgen und die Handlungsalternative mit dem jeweils größten erwarteten Nutzen wählen. In der RCT umfasst der Nutzenbegriff sowohl monetäre als auch nicht-monetäre Nutzen (z. B. die Befriedigung des Bedürfnisses nach sozialer Anerkennung). In diesem Kontext sei auf die Debatte zur Definition und zum Geltungsbereich der RCT verwiesen, die aktuell zwischen Karl-Dieter Opp und Hartmut Esser ausgetragen wird. Während Opp (2019) sämtliche Erweiterungen als kompatibel mit der RCT interpretiert und kaum empirische Phänomene außerhalb des Geltungsbereiches der RCT identifiziert, nimmt Esser (2018) zum Teil eine Gegenposition ein und weist auf Phänomene hin, die mit dem grundlegenden Optimierungsmechanismus der RCT selbst in ihren weitesten Varianten nicht mehr erklärbar sind.

1 Die gewählte männliche Form schließt eine adäquate Sprachform für weiblich und divers ein.

Ausgehend vom Paradigmenwechsel in der Ökonomie von der älteren, paretianischen Wohlfahrtsökonomik (Musgrave, 1959) hin zur Neuen Politischen Ökonomie (Buchanan & Tullock, 1962; Downs, 1957; Niskanen, 1971; Olson, 1965) wurden in der Folgezeit die Modellparameter von Nutzentheorien und der RCT schrittweise erweitert. Diese Erweiterungen dienen dazu, auch Empirie deutend erklären zu können, die engen RCT-Varianten zu widersprechen scheint. Hierzu zählen die Einbeziehung der subjektiven Perspektive (Edwards, 1954), *bounded rationality* (Simon, 1955), Framing-Phase und nicht-lineare Nutzen- und Erwartungsfunktionen (Prospect-Theorie [Kahneman & Tversky, 1979]; kumulative Prospect-Theorie [Tversky & Kahneman, 1992]), Dual-Process-Theorien (z. B. das Elaboration Likelihood Model [ELM] von Petty & Cacioppo [1986], das MODE-Modell von Fazio [1990], das Heuristisch-Systematische Modell [HSM] [Chaiken, 1980; Chaiken, Liberman, & Eagly, 1989]) bis hin zu uneigennütigen sozialen Präferenzen (siehe hierzu Fehr & Schmidt [2006] für einen Überblick) sowie antisozialen Präferenzen (Abbink & Sadrieh, 2009). Gemeinsam ist den weiten RCT-Varianten, dass sie die Modellparameter der Theorie variieren (Kosten, Nutzen, Route der Informationsverarbeitung), jedoch die Annahme eines Maximierungskalküls als grundlegenden Selektionsmechanismus beibehalten. Anomalien werden entweder durch Zuhilfenahme zusätzlicher Konstrukte relativ kleiner Reichweite erklärt (z. B. Biases, Heuristiken) oder als unsystematisches *Rauschen* interpretiert und daher nicht weiter beachtet. Aktuell wird jedoch in der wirtschafts- und sozialwissenschaftlichen Literatur zunehmend über Phänomene berichtet, die auch von sehr weiten RCT-Varianten nicht erklärt werden können, wie beispielsweise die hohe Designsensitivität in experimentellen Diktatorspielen (objektiv marginale Änderungen des Experimentaldesigns können starke Effekte haben) oder die Identifizierung immer neuer Verhaltenstypen in Öffentliches-Gut-Spielen (Engel, 2011; Fallucchi, Luccasen, & Turocy, 2019; Goerg, Rand, & Walkowitz, 2019).

Die empirisch-analytische Politikwissenschaft ist nicht nur auf der Mikroebene mit der Krise der RCT konfrontiert, sondern es stellt sich darüber hinaus die Frage, wie auf Basis von nicht zufriedenstellend modellierten Wirkmechanismen auf der Mikroebene anschließend Folgerungen für die Aggregatebene gezogen werden können. Der methodologische Individualismus fordert zwingend, dass zur Erklärung von kollektiven politischen Entscheidungen alle drei Glieder der Erklärungskette ineinandergreifen müssen: (a) Die subjektive Wahrnehmung des institutionellen Handlungsrahmens, auf der (b) die Handlungsselektion von politischen Akteuren beruht, sowie (c) deren Aggregation in Form der kollektiven Resultate politischer Entscheidungsprozesse und Wahlen.

Methodisch steht die empirisch-analytische Politikwissenschaft in einer Tradition, das Verhalten von Akteuren im politischen Kontext pragmatisch und lösungsorientiert zu erforschen. Sie bedient sich hierzu eines breiten Instrumenten- und Methodenrepertoires, beispielsweise experimenteller Studien zu Mehrheitsentscheidungen (Sauermann, 2018, 2020; Sauermann & Kaiser, 2010), Fallstudien (Blatter, Janning, & Wagemann, 2007; Jahn, 2013; Lijphart, 1975), oder der Method of Difference (Lijphart, 1971; Mill, 1843). Die empirisch-analytische Politikwissenschaft hat

die Fortentwicklung der RCT nicht nur begleitet, sondern eigene Impulse, beispielsweise im Bereich der handlungstheoretischen Fundierung in der experimentellen kooperativen Spieltheorie, gesetzt (siehe hierzu etwa Morton & Williams [2010] für die experimentelle Methode in den Politikwissenschaften). Andererseits spiegelt sich die Forschung zu Institutionen wiederum in der spieltheoretischen wirtschaftswissenschaftlichen Forschung wider, wie etwa bei Fochmann, Jahnke, und Wagener (2019), die experimentell die Wirkung von vertrauenswürdigen Institutionen auf das Beitragsverhalten im Öffentliches-Gut-Spiel (ÖGS) untersuchen. Zudem werden in der empirisch-analytischen Politikwissenschaft Konzepte aus alternativen theoretischen Paradigmen (z. B. Strukturfunktionalismus, Makrotheorien, [neo-]marxistische Ansätze, Symbolischer Interaktionismus) und benachbarten Fachdisziplinen (z. B. Soziologie, Psychologie, Ökonomie) genutzt, wenn dies dem Erkenntnisgewinn dient. So wurde beispielsweise von Hagauer, Kittel und Schwaninger (2019) das Modell der Frame-Selektion (MFS) (Esser, 1996, 2001a, 2001b, 2002a, 2002b, 2003, 2010a, 2010b, 2018, 2019; Kroneberg, 2005, 2011, 2014; Esser & Kroneberg, 2015) in die politikwissenschaftliche Diskussion eingeführt, das als General Theory of Action weiterhin von menschlicher Entscheidungsfreiheit und dem methodologischen Individualismus ausgeht, jedoch die Annahme eines Maximierungskalküls aufgibt zugunsten des Match als grundlegendem Selektionsmechanismus. In den Reizen und „signifikanten“ Symbolen (Cues) einer Situation werden hierbei von Individuen typische Muster wiedererkannt und spezifische (matchende) mentale Modelle aktiviert, die mit entsprechenden Verhaltensprogrammen (Skripte) verbunden sind. Der Grad des Match beschreibt die Stärke der Verbindung zwischen den subjektiv wahrgenommenen Stimuli und einem mentalen Modell. Passen Stimuli und mentales Modell perfekt und ohne Störung zusammen, wird das zugehörige Verhalten automatisch ausgeführt. Existiert in einer Situation kein passendes mentales Modell oder wird ein bestehendes erschüttert (Mis-Match), so kann es zu mentalen Abwägungsprozessen und heuristischen Entscheidungen bis hin zu RCT-kompatiblen Optimierungskalkülen als Spezialfall kommen. Im Widerspruch zur RCT kann der Match, also die Passung eines mentalen Modells zu einer Situation, zu dem Ergebnis führen, dass Individuen die (matchende) Alternative mit dem geringsten Nutzen wählen, z. B. einem unfreundlichen Kellner Trinkgeld geben, obwohl bereits vor der Entscheidung klar ist, dass sie sich stundenlang über das gegebene Trinkgeld ärgern werden. Ein anderes Beispiel wäre das automatisierte Ausweichen eines Autofahrers mit seinem KFZ in eine Leitplanke, um nicht einen Hasen auf der Fahrbahn zu überfahren. Das Ausmaß mit der RCT unvereinbarer Empirie und das Vorliegen alternativer theoretischer Erklärungsansätze geben den Anstoß für die vorliegende Dissertation. Anhand unterschiedlicher empirischer Messinstrumente auf verschiedenen Granularitätsebenen, von der Augenbewegungsmessung über individuelles Entscheidungsverhalten und das Wahlverhalten in großen Gruppen bis hin zur Sekundäranalyse einer repräsentativen Bevölkerungsumfrage wird die Erklärungskraft des Rational-Choice-Ansatzes gegenüber alternativen Verhaltenstheorien und Selektionsmechanismen untersucht.

Mikro-Makro-Debatte in der vergleichenden Politikwissenschaft

Zur Erklärung von Anomalien individuellen Verhaltens im institutionellen Handlungsrahmen werden in der Politikwissenschaft nicht nur individualistische, handlungstheoretische Ansätze wie die RCT und ihre oben skizzierten Erweiterungen diskutiert, sondern ergänzend wird auch auf makro-orientierte, institutionalistische oder systemtheoretische Forschungstraditionen rekurriert. Im Spannungsfeld zwischen Mikro-Ansätzen (Behaviorismus, ökonomische Handlungstheorien, Rational-Choice-Institutionalismus) und Makrotheorien (Klassifikation politischer Ordnungen, makro-zentrierte institutionalistische Ansätze, systemtheoretische Analyse) wird vor allem die Frage der Bedeutsamkeit des sozialen Feldes (z. B. Institutionen, Kultur, Normen) für das Verhalten (politischer) Akteure erörtert. Dies spiegelt sich auch in der Debatte über das Menschenbild wider, die ihren Ausgangspunkt in der Gegenüberstellung der Archetypen eines nutzenmaximierenden *Homo oeconomicus* und eines an Normen und Werten orientierten *Homo sociologicus* (Dahrendorf, 1958) hat. Im Kern steht hierbei die Frage, welchen direkten oder indirekten Einfluss Umwelt, Kultur und (politische) Institutionen auf Individuen ausüben.

Nicht nur in Bezug auf den Einfluss von Institutionen als unabhängige Variable (UV) auf das individuelle Verhalten als abhängige Variable (AV), sondern umgekehrt auch bezüglich des Einflusses von Verhalten (UV) auf Institutionen (AV) stehen sich die mikro-orientierte RCT und Makrotheorien als Extrempositionen gegenüber. Seit der klassischen Debatte zwischen den Ansätzen von Émile Durkheim und Max Weber (Ragin & Zaret, 1983) wird in den Sozialwissenschaften diskutiert, ob reine Makrotheorien zulässig sind oder ob es einer Mikrofundierung durch die Aggregation von Ergebnissen der unteren Ebenen (Meso- oder Mikroebene) bedarf. Ein Grund für die anhaltende Debatte zwischen Makro- und Mikro-Ansätzen zu der Frage, zu welchem Grad Institutionen das Verhalten von Akteuren bestimmen, liegt in dem Umstand, dass beide Seiten den zugrundeliegenden Mechanismus nicht zufriedenstellend erklären können, auf welche Weise der institutionelle Handlungsrahmen auf das individuelle Verhalten wirkt. Konkret: Wie gelangt eine Institution in das Bewusstsein und in das Handeln politischer Akteure?

Makro-orientierte systemtheoretische Ansätze sind im Gegensatz zu ihren eher hermeneutisch-beschreibend geprägten Vorläufern empirisch orientiert. Sie gehen davon aus, dass Individuen funktional im Sinne der Institution handeln und dasjenige Verhalten zeigen, das zur Institution und den entsprechenden Rollenerwartungen passt. Politische Institutionen müssen dafür sorgen, dass ihre Regeln ins Bewusstsein der Akteure gelangen. Tun sie dies nicht, so werden Institutionen dysfunktional und instabil, beispielsweise wenn sie den internen Zusammenhalt und die Inklusion ihrer Akteure nicht ausreichend herstellen. Integration stellt hierbei in Talcott Parsons' AGIL-Schema (Parsons, 1951) die dritte von vier Kernfunktionen eines Systems dar. Parsons postuliert zudem, dass individuelle Handlungen prinzipiell von Normen geprägt sind, und widerspricht damit den Annahmen des metho-

dologischen Individualismus (Thiery, 2019)². Mit anderen Worten: Entweder politische Institutionen sorgen dafür, dass sie funktional ins Bewusstsein von Akteuren gelangen, oder sie gehen unter. Gerade diese Frage der Erhaltung des politischen Systems steht bei Parsons sowie bei David Easton (1965) als weiterem Vertreter der Systemtheorie im Vordergrund.

Der Neue Institutionalismus (March & Olsen, 1983, 2006) beinhaltet in seinen Varianten sowohl Aspekte der Mikro- als auch der Makroebene. So lassen sich die Varianten des Neuen Institutionalismus entlang der Dimension aufspannen, zu welchem Grad eine Institution handlungsleitende Verhaltensmuster bereitstellt und deren Ausführung erzwingt (soziologisch-strukturtheoretisch verankerte Konzepte) oder, umgekehrt betrachtet, in welchem Ausmaß die Restriktionen und Beschränkungen von Institutionen in die subjektiven Nutzenkalküle der Individuen eingespeist werden (Rational-Choice-Institutionalismus). Im Spannungsfeld zwischen diesen beiden Richtungen befindet sich der Historische Institutionalismus, der sowohl die Kette prägender Weichenstellungen von Institutionen und ihre Wirkung auf individuelle Reaktions- und Anpassungsprozesse politischer Akteure betrachtet, als auch die kollektiven Ergebnisse im Aggregat sowie die Rückkopplungsprozesse auf die langfristige dynamische (pfadabhängige) Weiterentwicklung von Institutionen analysiert (Kaiser, 2001).

Der unter anderem von Dowding und King (1995) vertretene Rational-Choice-Institutionalismus sieht Institutionen als Gleichgewichtslösungen für das Koordinationsproblem bei öffentlichen Gütern. Ostrom (2007) geht auf das Menschenbild im Institutional Analysis and Development (IAD) Framework ein und weist auf die Vorteile hin, die es mit sich bringt, in vielen Kontexten die strikte Annahme eines Homo oeconomicus aufgrund ihrer mathematischen Formulierbarkeit und großen Leistungsfähigkeit (wie in der neoklassischen Wirtschaftswissenschaft bzw. der Spieltheorie modelliert) aus pragmatischen Gründen beizubehalten, auch wenn dieser enge Rational-Choice-Ansatz empirisch in vielen Anwendungsfeldern teilweise widerlegt wird. In diesem Zusammenhang weist Ostrom auch auf die von Simon (1955) untersuchten Aspekte der *bounded rationality* hin. Allerdings geht Ostrom davon aus, dass sich durch wiederholte Interaktionen das Problem der fehlenden Information (teilweise) beheben lässt (Ostrom, 2007, S. 31) – dies steht im Einklang mit empirischen Befunden aus der Verhaltensökonomie. So können Brosig-Koch, Riechmann und Weimann (2007) in ihren Experimenten durch wiederholte Interaktion von Individuen zunehmend eigennutzmaximierendes Verhalten nachweisen.

Der Historische Institutionalismus als ein alternativer Ansatz versteht sich als pfadabhängig und weicht in Teilen stark von der RCT ab (Peters, 2019). In Bezug auf das Menschenbild beschäftigt sich der Historische Institutionalismus hauptsächlich mit der Fragestellung, ob Präferenzen als endogener Faktor von Institutionen geformt werden (Thelen & Steinmo, 1992) oder exogen vorgegeben sind, wie dies vom Rational-Choice-Institutionalismus angenommen wird. Hierzu argumentiert

2 Die große Leistung dieses Paradigmas war Parsons' (1969) korrekte Prognose des Untergangs der realsozialistisch geprägten Wirtschaftssysteme.

Kaiser (1997), dass diese Unterschiedlichkeit bezüglich der Provenienz von Präferenzen keinen grundsätzlichen Unterschied der beiden Ansätze bedeute, sondern von der jeweiligen wissenschaftlichen Fragestellung abhängt. Thelen (1999) plädiert für eine gegenseitige Befruchtung von mikro-fundiertem Rational-Choice-Institutionalismus und dem eher makro-orientierten Historischen Institutionalismus, kritisiert jedoch aus Perspektive des strikt empirisch orientierten Historischen Institutionalismus die starke Theorielastigkeit des Rational-Choice-Institutionalismus (Thelen, 1999). Demgegenüber geht der makro-orientierte Soziologische Neoinstitutionalismus davon aus, dass individuelle Entscheidungen immer im Kontext von Institutionen stehen. Institutionen stellen demnach die eigentlichen Sinnzusammenhänge für das Handeln dar (Powell & DiMaggio, 1991). Ein hierzu inverser Pfad, der die subjektive Perzeptions-, Handlungs- und Anpassungsprozesse in den Vordergrund stellt, wird mit dem New Constructivist Institutionalism bzw. diskursiven Institutionalismus (Discursive Institutionalism) (Schmidt, 2008; 2011) beschritten. Hier wird die statische Analyseverfahren aller oben skizzierten Institutionalismus-Richtungen kritisiert und durch eine dynamische Sicht ersetzt, die von endogenen Veränderungen innerhalb von Institutionen ausgeht und von der *Theorie des kommunikativen Handelns* (Habermas, 1981) geprägt ist. Bell (2011) argumentiert, dass sich hierbei gewisse Überlappungen zwischen dieser diskursiven Sicht und dem Historischen Institutionalismus ergäben, da zumindest einer der Unterstränge des Historischen Institutionalismus auf die aktive Rolle von Individuen fokussiert, wobei die Individuen von den Institutionen geformt, aber nicht beherrscht würden. Auch Peters (2019) weist auf die dynamische Entwicklung im Feld des Neuen Institutionalismus hin und betont dabei die zunehmend wichtiger werdende Rolle von informellen Institutionen.

Die empirisch-analytische Politikwissenschaft ist von den aktuellen uneinheitlichen empirischen Befunden der Verhaltensforschung unmittelbar betroffen, die teilweise den Prognosen der RCT grundlegend widersprechen, andererseits jedoch auch zentrale Prognosen des Rational-Choice-Institutionalismus (experimentell) stützen. Alle oben skizzierten Ansätze basieren auf expliziten und impliziten Annahmen zur (subjektiven) Perception von Institutionen und den zugrunde liegenden Mechanismen der Handlungskontrolle, beispielsweise der intersituationalen sowie intertemporalen Stabilität von Präferenzen oder der Handlungssteuerung durch eine zentrale Exekutive. Ein Ziel der vorliegenden Dissertation ist daher, zu einem besseren Verständnis der grundlegenden Mechanismen der Wahrnehmung, Handlungsselektion und Aggregation beizutragen.

Verhaltenstheorien

In der oben skizzierten Mikro-Makro-Debatte argumentieren Vertreter von Makrotheorien, dass diese zumindest einen Teil der empirischen Befunde besser erklären könnten als mikro-orientierte Verhaltenstheorien. Unter Mikrotheorien werden in der Literatur häufig enge Varianten des ökonomischen Verhaltensparadigmas ver-

standen, ohne dessen Erweiterungen oder alternative, ebenfalls dem methodologischen Individualismus verpflichtete Verhaltenstheorien in den Vergleich mit einzubeziehen. Das ökonomische Verhaltensparadigma wurde jedoch im Laufe der Zeit in unterschiedlich engen und weiten RCT-Varianten ausdifferenziert, die aktuell in teils voneinander getrennten Entwicklungspfaden vorangetrieben werden. Trotz großer Vielfalt der Theorien ist bei vielen RCT-Varianten ein ähnliches Muster ihrer pfadabhängigen Fortentwicklung identifizierbar: Modellparameter werden verändert, der grundlegende Selektionsmechanismus der RCT sowie die Annahme einer zentralen Exekutive werden jedoch nicht in Frage gestellt. Kontrastierend hierzu postulieren soziologische und psychologische Rahmenmodelle sowie neurowissenschaftliche Konzepte die Möglichkeit alternativer Selektionsmechanismen.

Die fünf Studien der Dissertation haben unterschiedliche Schwerpunkte bezüglich der betrachteten Verhaltenstheorien und angewendeten Methode. Nachfolgend werden die unterschiedlichen Verhaltenstheorien kurz skizziert, um anschließend die einzelnen Studien der Dissertation insoweit zu verorten und in einen inneren Zusammenhang zu stellen, auf dem der gemeinsame Ertrag aller Studien basiert. Diese Gesamtschau führt Paradigmen aus unterschiedlichen Fachdisziplinen in einer gemeinsamen Übersicht zusammen und ergänzt diese um weitere, aktuelle Verhaltenstheorien.

Zur Einordnung und zum Vergleich der unterschiedlichen individualistischen Verhaltenstheorien wird als Rahmen das auf dem methodologischen Individualismus basierende Mikro-Makro-Erklärungsschema der empirisch-analytischen Soziologie (Abbildung 1) herangezogen, dessen grafische Darstellung von Esser unter dem Begriff der Colemanschen Badewanne (Coleman, 1986; Esser, 1993, S. 98; 1999) im deutschsprachigen Raum weite Verbreitung gefunden hat.

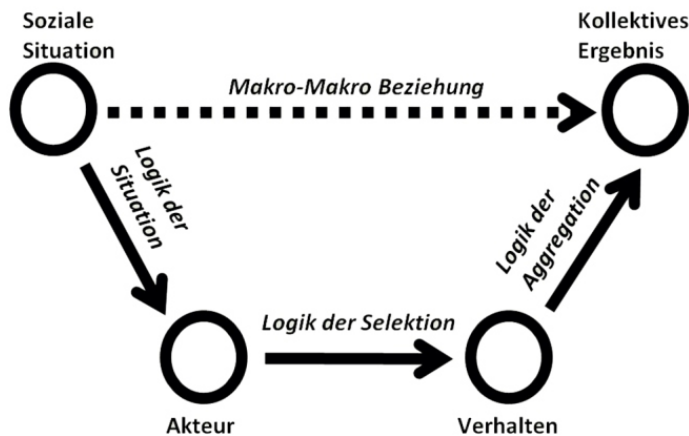


Abbildung 1 Soziologisches Mikro-Makro-Erklärungsschema (Colemansche Badewanne)

Das Mikro-Makro-Erklärungsschema soll nachfolgend am Beispiel von Kollektivgütern erklärt werden, die in drei Studien Untersuchungsgegenstand der Dissertation sind. Ausgangspunkt ist die Entscheidungssituation in Bezug auf ein öffentliches Gut (soziale Situation), die vom Akteur wahrgenommen (Logik der Situation) und bewertet wird und zu einem Verhalten führt (Logik der Selektion). Ein rationaler, eigennutzmaximierender Akteur sollte beispielsweise die Marginalität seines eigenen Beitrags erkennen und Trittbrettfahrerverhalten zeigen (Profitieren vom Nutzen des Kollektivgutes, ohne sich an den Kosten zu beteiligen). Verhalten sich viele Akteure als Trittbrettfahrer (Logik der Aggregation), kann aus der Aggregation individuell nutzenmaximaler Entscheidungen ein kollektives Ergebnis resultieren, bei dem sich alle schlechter stellen, indem das Kollektivgut nicht erstellt wird. Empirische Evidenz zeigt, dass die subjektiven Wahrnehmungs- und Entscheidungsprozesse nicht immer den Prognosen eines Homo oeconomicus folgen, sondern dass beispielsweise Framing, Salienz von Informationen (Fochmann & Wolf, 2019), periphere Informationsverarbeitungsprozesse und soziale Präferenzen von Bedeutung sind. Neben zweckrationalem kann auch wertrationales, affektuelles und traditionales (eingelebte Gewohnheit) Handeln im Sinne Max Webers (1921) auftreten. Abbildung 2 zeigt anhand von Piktogrammen der Colemanischen Badewanne, welche Prozesse auf welche Weise von den unterschiedlichen Verhaltenstheorien explizit modelliert werden. Die einheitliche Darstellungsform der heterogenen Ansätze soll einen Vergleich der Grundmechanismen der unterschiedlichen theoretischen Ansätze erleichtern.

Piktogramm (1) symbolisiert das standardökonomische Modell, die enge RCT1 und deren Erweiterungen zur RCT2 und zur RCT3 (Esser 2018, 2019). Der Kern der ökonomischen Nutzentheorie und der RCT ist die multiplikative Verknüpfung von Nutzen und Wahrscheinlichkeiten, die in den Sozialwissenschaften als Wert-Erwartungs-Theorien in den Begrifflichkeiten unterschiedlicher Kontexte ausformuliert werden. Beispiele sind hier Rotters Lerntheorie des sozialen Lernens (1954), die Einstellungstheorie von Fishbein und Ajzen (1974; 1975), der motivationstheoretische Ansatz von Vroom (1964), die Theorie der Rollenbilanz von Wiswede (1977) und Edwards' (1954) nutzentheoretisches *subjectively expected utility maximization model* (SEU-Theorie), das die subjektive Perspektive betont. In Piktogramm (1) wird durch den einzelnen schwarzen Pfeil veranschaulicht, dass diese mikro-basierten Theorien ausschließlich die Selektion des individuellen Verhaltens explizit modellieren und hierbei von rational-kalkulierenden Informationsverarbeitungsprozessen ausgehen.

Piktogramm (2): In den Wirtschaftswissenschaften wurde insbesondere durch die Prospect-Theorie (Kahneman & Tversky, 1979) die moderne Verhaltensökonomie begründet, die sich in der Folgezeit in ihren beiden Strängen Behavioral Economics und experimentelle Wirtschaftsforschung ausdifferenziert hat. In der Prospect-Theorie (Kahneman & Tversky, 1979) bzw. kumulativen Prospect-Theorie (Tversky & Kahneman, 1992) werden Entscheidungen als zweistufiger Prozess modelliert, der in Piktogramm (2) sowohl den schwarzen, gestrichelten (weil selektiv nur Aspekte der Situationslogik modellierenden) Pfeil zur Logik der Situation, als auch auf der zweiten Stufe den Pfeil zur Logik der Selektion umfasst. In der Editing-Phase erfas-

sen Individuen im Kollektivgutbeispiel die institutionelle Rahmung als Kontext. Im Gegensatz zu weiter unten skizzierten Theorien ist hierbei die Definition der Situation (passiv) von der objektiven Situation abhängig und wird somit nicht vom Individuum selektiert (Stocké, 2002). Die Spielsituation und das Framing sind somit unabhängige Variablen, die beispielsweise vom Experimentator induziert werden (Smith, 1976). In der an die Editing-Phase anschließenden Evaluationsphase werden Entscheidungen evaluiert und getroffen. Die Werte- und Wahrscheinlichkeitsfunktionen sind hierbei subjektiv formuliert und weichen erheblich von engen Varianten der RCT ab.

Piktogramm	Basiserklärungsmechanismen
(1) 	Standardökonomisches Modell, ökonomische Nutzentheorie (von Neumann & Morgenstern, 1944), Rational-Choice-Theorie (RCT1, RCT2, RCT3) (Esser, 2018, 2019; Opp, 2019), SEU-Theorie (Edwards, 1954)
(2) 	Zweistufige Modelle, z. B. Prospect-Theorie (Kahneman & Tversky, 1979), kumulative Prospect-Theorie (Tversky & Kahneman, 1992)
(3) 	Einstellungstheoretische Wert-Erwartungs-Theorien (Fishbein & Ajzen, 1974, 1975), Korrespondenzprinzip (Ajzen & Sexton, 1999), Konsistenzkontroverse (Schmitt, 1990)
(4) 	Dual-Process-Theorien z.B. Elaboration Likelihood Model (ELM) (Petty & Cacioppo, 1986), MODE-Modell (Fazio, 1990), Heuristisch-Systematisches Modell (HSM) (Chaiken, 1980; Chaiken, Liberman, & Eagly, 1989)
(5) 	Soziologie: Modell der Frame-Selektion (MFS) als zweistufige Dual-Process-Theorie (Esser, 1996, 2001a, 2001b, 2002a, 2002b, 2003, 2010a, 2010b, 2018, 2019; Kroneberg, 2005, 2011, 2014; Esser & Kroneberg, 2015)
(6) 	Neurowissenschaften: Inverse Kausalität von Entscheidung und Verhalten (Frith & Haggard, 2018; Haggard & Eimer, 1999; Kornhuber & Deecke, 1965; Libet, Gleason, Wright, & Pearl, 1983; Nann, Cohen, Deecke, & Soekadar, 2019)
(7) 	Psychologie: BRAC Rahmenmodell (Binding and Retrieval in Action Control) (Frings et al., 2020a; 2020b)

Abbildung 2 Schematischer Vergleich der Grundmechanismen ausgewählter individualistischer Verhaltenstheorien

Piktogramm (3) stellt das Korrespondenzprinzip (Ajzen & Sexton, 1999) und die Konsistenzkontroverse (Schmitt, 1990) dar, die auf einstellungstheoretisch formulierten Wert-Erwartungs-Theorien fußen. Die beiden gestrichelten Pfeile, die ungerichtet nach oben zeigen, symbolisieren allgemeine Einstellungen. Demgegenüber repräsentiert der erste kurze schwarze Pfeil, der in der Logik der Selektion in Richtung Verhalten zeigt, eine spezifische Einstellung zum Verhalten, z. B. im Sinne der Wert-Erwartungs-Theorie von Fishbein und Ajzen (1974; 1975). Der daran anschließende kurze schwarze Pfeil stellt die Verhaltensabsicht (Behavioral Intention = BI) dar und der dritte Pfeil den Prozess der tatsächlichen Verhaltensausführung (Behavior = B). Das Korrespondenzprinzip betont, dass eine geringe Korrelation zwischen Einstellungen und Verhalten häufig darauf zurückzuführen ist, dass Einstellungen nicht auf ein Verhalten bezogen sind. Dies wird im Piktogramm (3) durch die ungerichteten gestrichelten Pfeile symbolisiert. Zudem besteht auch zwischen Verhaltenseinstellungen, Verhaltensabsicht und tatsächlichem Verhalten ein geringer Zusammenhang. Dies konnte von Six und Eckes (1996) in einer Metastudie empirisch belegt werden und wird im Piktogramm (3) durch drei horizontale, aufeinanderfolgende Pfeile symbolisiert. Die Verhaltensökonomie ist zunehmend mit dieser Problematik konfrontiert, da spieltheoretische Verhaltensmessungen vermehrt von Einstellungserhebungen begleitet werden, um die Gründe der gewählten Strategien zu messen.

Piktogramm (4): Aus der Erkenntnis heraus, dass Menschen nicht über jede Entscheidung extensiv nachdenken, sowie mit dem Argument, dass extensive Entscheidungsprozesse sowohl zeit- als auch kostenaufwändig sein können (Kosten der Informationssuche, Kosten der Informationsverarbeitung, aber auch beschränkte Kapazitäten des menschlichen Gehirns), führte Simon (1955) das Konzept der *bounded rationality* ein, das in Ökonomie und Sozialwissenschaften weite Verbreitung gefunden hat. Dieses Konzept spiegelt sich in der Politikwissenschaft beispielsweise bei Downs (1957) wider, der die Theorie des rationalen Wählerverhaltens modelliert. *Bounded rationality* kann als eine der Wurzeln von Dual-Process-Theorien gesehen werden, die von zwei möglichen Routen der Informationsverarbeitung – einer zentralen, rational-kalkulierenden und einer peripheren, automatisch-spontanen – ausgehen. In Piktogramm (4) sind die zwei möglichen Routen der Informationsverarbeitung als parallele, schwarze Pfeile von der Logik der Situation aus symbolisiert, die beide auf das Verhalten gerichtet sind. In Dual-Process-Theorien wird nicht von einer reduzierten Informationsverarbeitung, sondern von einer andersartigen, peripheren Informationsverarbeitung ausgegangen. Gigerenzer und Todd (1999) weisen in diesem Zusammenhang mit ihrem Konzept der *Smart Heuristics* darauf hin, dass die Route der Informationsverarbeitung in keiner Weise eine Aussage über die Qualität der Informationsverarbeitung oder die Qualität ihrer Ergebnisse sei. Fortentwicklungen wie das weiter unten dargestellte MFS von Hartmut Esser oder das Unimodell der Überzeugung (Erb et al., 2003; Kruglanski, Pierro, Mannetti, Erb, & Chun, 2007) betonen zudem den fließenden Übergang zwischen den beiden Informationsverarbeitungsrouten.

Das Modell der Frame-Selektion (MFS) (Piktogramm 5) ist eine zweistufige Dual-Process-Theorie. Auf beiden Stufen können sowohl zentrale als auch periphere

Informationsverarbeitungsprozesse stattfinden. Im reflexiv-kalkulierenden *rc*-Modus als zentraler Route der Informationsverarbeitung werden die Informationen und Folgen einer Situation systematisch berücksichtigt. Demgegenüber wird im automatisch-spontanen *as*-Modus als peripherer Route der Informationsverarbeitung eine Situation heuristisch und auf Basis mentaler Modelle wahrgenommen (Kroneberg, 2005). Da sich in der aktuellen Version des MFS alle Prozesse auf zwei Routen vollziehen können, sind im Piktogramm (5) für die drei Prozessstufen jeweils zwei parallele schwarze Pfeile abgebildet: Für das Framing (Logik der Situation), für die Skriptselektion („Programm des Handelns“) und für die anschließende Handlungsselektion³ („Handeln“). Das MFS basiert auf Annahmen der SEU-Theorie (Edwards, 1954) und ist konsistent zu anderen Dual-Process-Theorien, die weiter oben skizziert wurden. Jedoch ist im MFS die Wahrnehmung der Situation die abhängige Variable, und das Framing kann ein mehr oder weniger aktiver gesteuerter, rational-kalkulierender Prozess sein. Mit dem *Match* wird im MFS zudem ein Selektionsmechanismus für mentale Elemente formuliert, der sich vom Nutzenmaximierungsprinzip unterscheidet.

Jenseits des RCT-Paradigmas und seiner Erweiterungen hat sich unter anderem in der Soziologie, der (Sozial-)Psychologie, der Wahrnehmungspsychologie und den Neurowissenschaften eine Vielzahl von (ebenfalls individualistischen, mikro-orientierten) Konzepten zu Aspekten der Handlungssteuerung etabliert, die sich grundlegend von den Annahmen des RCT-Universums unterscheiden.

Piktogramm (6) stellt den neurologischen Grundmechanismus dar, der in Folge der Entdeckung des Bereitschaftspotentials (Kornhuber & Deecke, 1965) robust nachgewiesen werden konnte. Zeitlich erst nach der Einleitung der Verhaltensausführung (gemessen als spezifisches oder unspezifisches Bereitschaftspotential im motorischen Kortex) wird diese Information der Handlungseinleitung im Bewusstsein integriert und der bewusste Wille (*urges*) zur Ausführung der Handlung empfunden. In Piktogramm (6) wird durch den schwarzen, vom Verhalten zum Akteur rückwärtsgerichteten, gestrichelten Pfeil symbolisiert, dass das Elektroenzephalogramm-Signal der Änderung des (unspezifischen) Bereitschaftspotentials bereits bis zu zwei Sekunden vor der Handlungsausführung gemessen werden kann, der Wille jedoch erst unmittelbar vor der gemessenen Handlungsausführung empfunden wird (Frith & Haggard, 2018; Haggard & Eimer, 1999; Libet, Gleason, Wright, & Pearl, 1983; Nann, Cohen, Deecke, & Soekadar, 2019). In Bezug auf subjektiv empfundene und objektiv gemessene Zeitpunkte ist anzumerken, dass das Gehirn keine objektive Zeit kennt, sondern sowohl Gleichzeitigkeit als auch zeitliche Abfolgen eine subjektive Wahrnehmung darstellen. Die inverse Kausalität von Verhaltensentscheidung und Verhaltensausführung hat unter anderem in der Rechtswissenschaft

3 Zur Vermeidung von Missverständnissen ist anzumerken, dass in der grafischen Darstellung von Esser und Kroneberg (2015, S. 71) die Handlungsselektion als aufwärtsgerichteter Pfeil abgebildet ist, da Verhalten („Action A_k “) an der Position des kollektiven Ergebnisses abgebildet ist.

und Philosophie zu einer anhalten Debatte über die Willensfreiheit des Menschen geführt.

Untersuchungsgegenstand des in Piktogramm (7) dargestellten BRAC-Rahmenmodells (Binding and Retrieval in Action Control) (Frings et al., 2020a; 2020b) sind Episoden, die als mentale Repräsentation von Reizen, Reaktionen und Konsequenzen eine zusammengefügte Einheit bilden. Der Erwartungsnutzen im Kontext des BRAC-Rahmenmodells stellt hierbei ein attribuiertes Merkmal der für die Zukunft erwarteten Konsequenzen dar. Die drei auf den Akteur gerichteten Pfeile aus Richtung der Situation (Reize), des Verhaltens (Reaktion) und der Ergebnisse (Konsequenzen) spiegeln die Integration der entsprechenden Merkmale und Attribute in die jeweilige Episode wider. Die in umgekehrter Richtung zeigenden Pfeile symbolisieren den Abruf von Episoden sowie deren Wirkungen auf das beobachtbare Handeln.

Forschungsfrage

Ein Grund für die Zunahme empirischer Evidenz, die auch von weiten Varianten des RCT-Paradigmas nicht mehr erklärt werden kann, könnte in der zunehmenden Verbreitung neuer Forschungsmethoden liegen, beispielsweise spieltheoretischer Experimente, faktorieller Surveys oder bildgebender Verfahren in der Medizin. In der Literatur (siehe dazu beispielsweise Manzo [2013]) wird diskutiert, ob (1) die RCT beibehalten werden kann (z. B. Opp, 1978; 2019), (2) die RCT im Sinne eines „neoclassical repairshop“ (Brandstätter & Güth, 1994) noch stärker zu erweitern ist, (3) der Grundmechanismus (Selektion der nutzenmaximalen Alternative durch eine zentrale Exekutive) aufgegeben werden muss (siehe hierzu beispielsweise Braun und Wolbring [2012]) oder (4) eine theorielose Variablen-Forschung betrieben werden soll, deren Ergebnisse von ökonometrischen Analyseinstrumenten und verfügbarer Datengrundlage determiniert wird. Um einen Beitrag zu dieser Diskussion zu leisten, wird in der vorliegenden Dissertation menschliches Verhalten in verschiedenen Anwendungsfeldern auf unterschiedlicher Granularität untersucht, von basalen Augenbewegungen über subjektive Gründe für individuelles Entscheidungsverhalten bis hin zu politischen Einstellungen bezüglich Finanzierungsoptionen im Gesundheitswesen in einer repräsentativen Bevölkerungsstudie. Unter Verwendung eines Methodenmix zur Messung der mentalen Prozesse wird untersucht, wie groß das Ausmaß der Abweichungen von Prognosen der RCT werden kann und ob alternative Mechanismen der Handlungssteuerung eine größere Vorhersagekraft besitzen. Insbesondere wird hierzu das MFS betrachtet, das seine Wurzeln sowohl in Parsons *unit act* (1937) als auch in der RCT hat. Dieser Umstand der Verwurzelung des MFS sowohl in der Mikro- als auch in der Makroebene, könnte dafür sprechen, dass das MFS ein bisher übersehenes Bindeglied zwischen institutionalistischen Makrotheorien und RCT-orientierten Mikrotheorien darstellt, indem es die Wirkmechanismen offenlegt, wie der institutionelle Rahmen über Cues und die Aktivierung mentaler Modelle in das Bewusstsein der politischen Akteure gelangt und ihre Entscheidungen beeinflusst.

2 Beitrag der Dissertation zur forschungstheoretischen Diskussion

Alle Beiträge der Dissertation folgen der übergeordneten Überlegung, anhand unterschiedlicher Methoden die Leistungsfähigkeit verschiedener verhaltenstheoretischer Ansätze zu untersuchen, um individuelles und kollektives Verhalten in unterschiedlichen Anwendungsbereichen zu erklären. Insbesondere wird die Frage untersucht, ob die grundlegende Annahme der Selektion der nutzenmaximalen Handlungsoption durch eine zentrale Exekutive eine ausreichend große Erklärungskraft aufweist oder ob alternative Selektionsmechanismen wie beispielsweise der *Match* mentaler Bewusstseinsselemente im Sinne des Modells der Frame-Selektion (MFS) ein leistungsfähigeres Konstrukt darstellen, um die empirische Evidenz zu erklären und um den Mechanismus offenzulegen, wie beispielsweise der institutionelle Handlungsrahmen in das Bewusstsein von (politischen) Akteuren gelangt und ihre Entscheidungen beeinflusst.

Einzelne Beiträge der Dissertation und individueller Beitrag des Autors

Den Anfang der vorliegenden Dissertation bildet Studie 1 zum Öffentliches-Gut-Spiel, da dieses den zentralen theoretischen und methodischen Kern zum Verständnis der anderen Studien bildet und den Rahmen aufspannt, in dem diese angesiedelt sind. Die anschließend dargestellte Studie zu Augenbewegungen (Studie 2) betrachtet die grundlegenden Mechanismen auf kleinster Granularität und zeigt die Bandbreite von Verhalten (bewusstes und unwillkürliches Augenbewegungsverhalten) auf. Das Voting-Experiment in Studie 3 ist so angelegt, dass gezielt Trittbrettfahrerverhalten eliminiert wird, und stellt einen ersten, noch relativ abstrakten Anwendungsfall von Verhaltenstheorien dar. Der Beitrag zu Finanzierungsoptionen in der Krankenversicherung (Studie 4) kombiniert mehrere Sekundäranalysen und fächert Methoden für diesen inhaltlichen Anwendungsfall auf. Damit wird die Basis gelegt für den Artikel zum Diktatorspiel (Studie 5), dessen Befunde auf Aggregatebene im Kontext der anderen Studien plausibel werden.

(Studie 1) In der Untersuchung *Konditionale Kooperation in experimentellen Öffentliches-Gut-Spielen mit und ohne Feedback: Erklärung oder Artefakt?* (Betz, 2020) werden in einem ausführlichen Einführungsteil die methodischen und inhaltlichen Annahmen sowie die Wirkung unterschiedlicher Faktoren (z. B. limitiertes Feedback) auf wiederholte ÖGS diskutiert. Im empirischen Teil ($N = 32$) stellt das erste Treatment *No-Punishment* ($n = 16$) eine exakte Replikation des Öffentliches-Gut-Spiels ohne Bestrafung von Fehr und Gächter (2000) dar, das insgesamt dreimal hintereinander mit den gleichen Versuchspersonen (Vpn) jeweils über zehn Runden durchgeführt wird. In der ersten Runde bestehen keine signifikanten Unterschiede zu den

Daten von Fehr und Gächter (2000). Im weiteren Verlauf der Spielrunden zeigt sich zunehmend die Bedeutung des gruppeninternen, kollektiven Beitragsverhaltens für das Beitragsverhalten der einzelnen Spieler. Dies führt dazu, dass in zwei der vier Gruppen fast in jeder Runde der Höchstbetrag zum öffentlichen Gut beigetragen wird. Die Ergebnisse sind teilweise nicht mit der RCT erklärbar und zeigen die Bedeutung der Gruppenbeitragsverläufe für die individuellen Beiträge ab der zweiten Runde. In Abhängigkeit von den Gruppenbeitragsverläufen werden bei den Vpn völlig unterschiedliche Erlebniswelten, Heuristiken und Gerechtigkeitsprinzipien aktiviert, die sich aufgrund äußerer Reize und innerer Prozesse dynamisch wandeln können (Reframing). Derselbe Spieler kann dabei zu Beginn ein anderer Spielertyp sein als am Ende der Experimente. Eine in der Literatur bisher übliche Aggregation von Gruppen mit unterschiedlichen Verläufen ist daher nicht nur aus methodologischen, sondern auch aus empirischen Gründen unzulässig. Im zweiten Treatment *No-Feedback* ($n = 16$) wird den Spielern kein Feedback über die Beiträge der anderen Gruppenmitglieder gegeben. Entgegen der Prognosen der RCT tragen Spieler hier auch dann zum Gemeinschaftsgut bei, wenn weder die individuellen noch die kollektiven Auszahlungen mitgeteilt werden und somit auch keine negativen Konsequenzen für das eigene Trittbrettfahrerverhalten zu befürchten sind. Dies stützt das Konzept des *Match* im Modell der Frame-Selektion. Aus dem Konzept folgt, dass die anfänglichen Entscheidungsmuster beibehalten werden, wenn kein Anlass für eine Änderung beispielsweise durch Störreize besteht. Die von den Vpn geäußerten Entscheidungsgründe stützen diesen Befund.

(Studie 2) Der Beitrag *Fixation Durations – Why Are They So Highly Variable* (Galley, Betz & Biniossek, 2015) stellt die Frage, ob ultrakurze und ultralange Fixationsdauern des Auges systematisch auftreten und eigenständige Fixationsdauerklassen darstellen. Anhand einer Sekundäranalyse von Daten, die mithilfe einer Elektrookulografie¹ (EOG) erhoben wurden, wird in sechs unterschiedlichen Aufgaben (z. B. *Task Springender Punkt*, *Task Bildbetrachtung*, *Task Raven*) das Augenbewegungsverhalten gemessen ($N = 1131$). Hierbei werden unseres Wissens erstmals neben den bekannten kurzen und langen Fixationsdauern auch ultrakurze und ultralange Fixationsdauern nachgewiesen und deren Funktionen spezifiziert. Die besondere wissenschaftliche Bedeutung der Ergebnisse ergibt sich aus dem Umstand, dass alle vier Typen der Fixationsdauer sowohl bei Verbindung von Auge und Bewusstsein (*connected mode*) als auch bei voneinander entkoppelter, autonomer Funktion beider Systeme (*non-connected mode*), bei der sich die Okulomotorik in einer permanenten Orientierungsfunktion befindet, nachweisbar sind (z. B. im *Task Eyes Closed*). Dies ist ein weiteres Indiz für die Annahme einiger psychologischer Konzepte, dass (sensomotorisches) Verhalten nicht zwangsläufig einer zentralen Exekutive bedarf, wie dies von allen RCT-Varianten bisher behauptet wird.

(Studie 3) Die Untersuchung *Der Einfluss unterschiedlicher Gerechtigkeitsprinzipien auf Mehrheitsentscheidungen in einem spieltheoretischen Social-Good-Experiment*

1 Die Elektrookulografie misst im Gegensatz zu Eye-Tracking auch bei geschlossenen Lidern.

(Biniossek & Betz, 2018) erhebt anhand eines Voting-Experimentes ($N = 261$) den Einfluss unterschiedlicher sozialer Präferenzen und Gerechtigkeitsprinzipien auf Mehrheitsentscheidungen. In allen neun Treatments können Gruppen zwischen einer relativ gleichmäßigen Verteilung und einer relativ ungleichmäßigen Verteilung einer Geldsumme zwischen den Mitgliedern der Gruppe wählen, jedoch wird der Anteil der Benachteiligten, das Ausmaß der Benachteiligung sowie die Gruppeneffizienz zwischen den Treatments variiert. In acht der neun Treatments werden weite Rational-Choice-Ansätze auf Aggregatebene bestätigt, nämlich dass Individuen in bestimmtem Ausmaß soziale Präferenzen verfolgen, wenn dies wenig kostet (Low-Cost-Hypothese [Best & Kroneberg, 2012]), sich jedoch bei höheren Kosten Eigennutzmaximierung als dominante Strategie durchsetzt. Zwar erweist sich in acht der neun Treatments die Maximierung der eigenen Auszahlungen als stärkster Faktor, jedoch spielen möglichst hohe Auszahlungen an andere ebenfalls eine wichtige Rolle. Dies wird experimentell unter Konstanthaltung des GINI-Koeffizienten nachgewiesen. Jedoch werden in einem Treatment Motive wie Eigennutzmaximierung oder Ungleichheitsaversion durch den *Identifiable-Victim-Effekt* (Jenni & Loewenstein, 1997) als dominantes mentales Modell verdrängt. In diesem Treatment wird eine kleine Minderheit sehr stark durch eine negative Auszahlung benachteiligt. Diese Ergebnisse widersprechen Theorien der Ungleichheitsaversion (z. B. Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999), die als Erweiterung der RCT interpretierbar sind.

(Studie 4) Der Artikel *Fairness ohne Eigennutz. Handlungsannahmen in der ökonomischen Theorie und Auswirkungen auf Reformoptionen im Gesundheitswesen* (Biniossek, Betz, Fetchenhauer, & Lungen, 2007) thematisiert politische Präferenzen zu Finanzierungsoptionen der gesetzlichen Krankenversicherung und weist anhand einer Sekundäranalyse ($N = 1279$) (Christoph & Ullrich, 2006) nach, dass pauschalierte Krankenversicherungsbeiträge von Wählern aller Parteien mehrheitlich abgelehnt werden, am stärksten jedoch von Wählern von Bündnis 90/Die Grünen, die als Anhänger der „Partei der Besserverdienenden“ (Klein & Falter, 2003) stark von Kopfpauschalen profitieren würden. Anhand einer weiteren Sekundäranalyse einer repräsentativen Bevölkerungsumfrage ($N = 440$) (Pappi & Shikano, 2004) wird nachgewiesen, dass eine einkommensbezogene Finanzierung unabhängig vom individuellen Eigennutz stark präferiert wird, selbst wenn damit individuell erhebliche finanzielle Nachteile verbunden sind. Im Anwendungsfeld von Finanzierungsoptionen der gesetzlichen Krankenversicherung scheint das gedankliche Modell von *Fairness ohne Eigennutz* passend zu sein, also zu *matchen* und das Entscheidungsverhalten zu leiten. Dies widerspricht der RCT auch in seinen weiten Formen, da diese lediglich von Fairness aus Eigennutz ausgeht.

(Studie 5) Das Manuskript *The Squeaky Wheel Gets the Grease – Reference Point Variations in Game Theoretic Experimental Dictator Games* (Betz, Oestreich, & Biniossek, 2020) untersucht anhand von drei Varianten des experimentellen Diktatorspiels ($N = 72$), ob Individuen der vom Computer (Treatment *Referenzpunktvariation*) oder vom Rezipienten im Strategiemodus vorgegebenen Aufteilung (Treatment *Inverses Diktatorspiel*) folgen. Diktatoren transferieren von den verfügbaren 10 EUR durch-

schnittlich 5 EUR, wenn der Rezipient eine gleichmäßige Aufteilung der 10 EUR vorschlägt, und 3.42 EUR, wenn der Computer dies tut. Dagegen transferieren Diktatoren durchschnittlich 5.33 EUR, wenn der Rezipient 9 EUR auf dem Rezipientenkonto und 1 EUR auf dem Diktatorkonto vorgibt, und 3.75 EUR, wenn der Computer dies tut. Diese Ergebnisse widersprechen weiten Varianten der RCT und liegen signifikant über dem Benchmarkwert von Forsythe, Horowitz, Savin, und Sefton (1994).

Der individuelle Beitrag des Autors zu den fünf Studien ist Folgender: Studie 1 wurde in alleiniger Autorenschaft erstellt. Der individuelle Beitrag des Autors zu Studie 2 umfasst die Beteiligung an der Konzeption insbesondere des theoretischen Rahmens, an der Analyse und an der Ergebnisinterpretation sowie die Erstellung der Endfassung und die Übernahme der Kommunikation als korrespondierender Autor. Der individuelle Beitrag zu Studie 3 umfasst die Konzeption aus sozialwissenschaftlicher Perspektive, den Entwurf und die Berechnung des Gini-Koeffizienten sowie die Erstellung des hierauf basierenden Experimentaldesigns, die Durchführung der Experimente, die Auswertung der Daten, das Verfassen des Manuskripts und die Erstellung der Endfassung. Zu Studie 4 hat der Autor wesentliche Teile der Sekundäranalyse und Interpretation sowohl der repräsentativen Umfragedaten als auch der Experimentaldaten beigetragen sowie das Manuskript und die Endfassung maßgeblich mit erstellt. Als korrespondierender Erstautor von Studie 5 trägt der Autor die Gesamtverantwortung und hat den Hauptanteil der Leistung in allen Teilaspekten erbracht.

Gemeinsamer Ertrag der Studien

Nachfolgend wird thematisch untergliedert der Beitrag der einzelnen Studien zur Beantwortung der übergeordneten Forschungsfrage des individuellen Verhaltens im sozialen Feld (z. B. institutioneller Handlungsrahmen) skizziert und in den Gesamtrahmen der Dissertation eingeordnet, um sowohl den einzelnen als auch den gemeinsamen Ertrag der Studien aufzuzeigen. Den Schwerpunkt bilden die Analyse und die empirische Überprüfung der grundlegenden Mechanismen von Verhaltenstheorien, von hypothetischen Konstrukten mittlerer oder kleiner Reichweite sowie von deren Beziehung zueinander. Ein Bindeglied zwischen den heterogenen Studien ist der Bezug auf die unterschiedliche Granularität von Theorien, deren Verortung in der experimentellen Forschung bei unterschiedlichen Erhebungsgegenständen sowie die analytische Betrachtung der Ergebnisse sowohl auf Individual- bzw. Mikroebene als auch auf Aggregat- bzw. Makroebene.

Ein gemeinsames Ziel der Einzelstudien der Dissertation ist es, zur Erforschung der übergeordneten Fragestellung Theorien und Methoden aus unterschiedlichen Forschungstraditionen und Fachdisziplinen zusammenzubringen und anhand des gezielt entwickelten Experimentaldesigns sowohl Schwächen als auch Stärken der RCT insbesondere auf kleiner Granularität aufzuzeigen. Gleichzeitig wird diskutiert, in welchen Zusammenhängen alternative Ansätze wie das MFS oder das BRAC-Rahmenmodell einen zusätzlichen Erklärungsbeitrag leisten können. Vier der fünf Stu-

dien stützen sich auf die Methode der experimentellen Spieltheorie, die teilweise mit weiteren Messinstrumenten (beispielsweise aus der empirischen Umfrageforschung) kombiniert wird. Nachfolgend werden die Stärken und Schwächen der RCT im Vergleich zu alternativen Konzepten dargestellt und der gemeinsame Ertrag der Studien aufgezeigt. Abschließend werden methodische Implikationen und Limitationen der Dissertation skizziert.

RCT

Aus dem Menschenbild des Homo oeconomicus folgt, dass jedes Individuum ausschließlich eigennutzmaximierend handelt. Überprüft werden mit den Studien 1, 4 und 5 die auf diesem Menschenbild aufbauenden Hypothesen, dass in anonymen Experimenten keine Beiträge zu Kollektivgütern geleistet werden (ÖGS – Studie 1), an Rezipienten kein Geld verschenkt wird (Diktatorspiel – Studie 5) und bei Abstimmungen für die ökonomisch vorteilhafteste Option oder Partei votiert wird (Finanzierungsoptionen im Gesundheitswesen – Studie 4). Empirisch widerlegen die Studien der Dissertation massiv und grundlegend diese Annahmen und Prognosen der RCT. Nachfolgend soll daher diskutiert werden, inwiefern auch gegen die oben skizzierten weiten Varianten der RCT (Piktogramme 1 bis 4) grundlegende Verstöße vorliegen.

Während Studie 2 die RCT-Annahme einer zentralen Exekutive sowie analoger Informationsverarbeitungsprozesse schwächt, sind in den anderen vier Studien Verhaltensentscheidungen zu beobachten, die nicht durch die von der RCT angenommenen Mechanismen erklärbar sind. In beiden Treatments zum ÖGS werden zu hohe Beiträge geleistet, durch die RCT unerklärte Beitragsmuster gezeigt und inkonsistente Verhaltensgründe genannt. In Studie 3 sind in allen Treatments Entscheidungen beobachtbar, die bezogen auf Gerechtigkeitsmaße (Gini-Koeffizient) nicht zweckrational sind. Studie 4 zu Krankenversicherungsbeiträgen weist in einer repräsentativen Bevölkerungsumfrage unerwartete prosoziale Einstellungen nach. Auch in Studie 5 sind hohe Beiträge zu beobachten, zudem reagieren die Vpn stark auf Designvariationen. Beide Befunde dürften gemäß der RCT nicht auftreten. In jeder der fünf Studien werden somit empirische Ergebnisse präsentiert, die sogar weite Versionen der RCT (Piktogramme 1 bis 4) schwächen.

Neben dieser allgemeinen Schwächung der RCT werden in einigen Treatments zudem grundlegende Verstöße gegen die RCT und ihre postulierten Mechanismen nachgewiesen. Vier der fünf Studien widersprechen der RCT grundlegend: In Studie 1 werden im Treatment mit Feedback in zwei der vier Treatment-Gruppen unkonditional maximale Beiträge zum Gemeinschaftsgut geleistet. Von diesen Gruppen ist eine sogar zunächst von einem geringeren Beitragsniveau gestartet und verstößt daher massiv und elementar gegen die RCT, wie auch zwei Drittel der Vpn im Treatment, in dem die Spieler kein Feedback erhalten. In Studie 2 (Augenbewegungen) „verhält“ sich das Auge auch dann, wenn das Bewusstsein anderweitig beschäftigt ist (*non-connected mode*), und zeigt dabei die gleichen vier Typen der Fixationsdauer wie im *connected mode*. Im Voting-Experiment (Studie 3) verdrängt der *Identifiable*

Victim-Effekt in einem der neun Treatments sowohl Eigennutz als auch soziale Präferenzen und dominiert damit bei einigen Vpn die RCT-kompatiblen Strategien. In Studie 5 (IDG-Treatment) wird entgegen der RCT zu 100 % die Gleichverteilung (equal split) realisiert. Anhand der Strategiemethode konnten zudem bei vier der 12 Vpn in der Rolle des Diktators Beiträge oberhalb des equal split nachgewiesen werden, die von Theorien der Ungleichheitsaversion (ERC-Modell von Bolton & Ockenfels [2000]; Fehr & Schmidt Modell [1999]) kategorisch ausgeschlossen werden und grundlegend gegen die RCT verstoßen. Auch in Studie 4 zu Finanzierungsoptionen im Gesundheitswesen bevorzugten viele Besserverdienende die für sie finanziell nachteilige Variante solidarischer Beiträge. Selbst wenn die RCT soweit gedehnt wird, dass auch rationale Altruisten und Heilige erfasst werden, besteht widersprechende Evidenz: Da Teilnehmer von Experimenten (z. B. Rezipienten in Studie 5) in der Regel weniger bedürftig sind als beispielsweise notleidende Menschen in Krisengebieten, sollte rationale Wohltätigkeit dazu führen, nach dem Experiment einen Teil des verdienten Geldes an Bedürftige zu spenden. Dies wird im Voting-Experiment vor allem von einigen im Spiel als „Egoisten“ klassifizierten Spielern getan. Dieser intrapersonale Verhaltenswechsel widerspricht der RCT. Auch der *Identifiable-Victim-Effekt* widerspricht der RCT, da er sowohl eigennuttschädlich als auch gesamtgruppenschädigend zugunsten einer sehr kleinen Minderheit wirkt.

Als gemeinsamer Ertrag der Studien kann damit zwar einerseits eine große Leistungsfähigkeit der RCT nachgewiesen werden, jedoch wird andererseits den Grundmechanismen und Kernaussagen (Nutzenmaximierungsprinzip) sowie den daraus abgeleiteten Konstrukten (Verhaltenstypen) widersprochen, was zu der Schlussfolgerung führt, dass nicht mit der Theorie kompatible Ergebnisse mehr als nur ein *Rauschen* darstellen. In der Gesamtschau ergibt sich, dass die RCT auch bei Erweiterung um soziale Präferenzen in den hier präsentierten Studien geschwächt wird.

Ist die RCT am Ende?

Die RCT wird in den Studien der Dissertation nicht nur geschwächt, sondern anhand einer Vielzahl von Befunden auch gestützt. In der Gesamtschau der Studien wird sichtbar, dass Abweichungen von den Prognosen der RCT nicht zufällig, sondern systematisch auftreten. Die Frage ist somit nicht, welche Zukunft die RCT aufgrund ihrer Schwächung hat, sondern welche Dimensionen identifiziert werden können, in denen sich die RCT auf der einen Seite und alternative Konzepte auf der anderen Seite einer Skala der Erklärungsstärke gegenüberstehen. Diese Dimensionen werden nachfolgend dargestellt.

High-Cost versus Low-Cost

Vor allem in der Soziologie (Best & Kroneberg, 2012; Diekmann & Preisendörfer, 1992) wird die Diskussion geführt, ob prosoziales Verhalten nicht vor allem in Situationen auftritt, in denen es wenig kostet. Isoliert betrachtet zeigt sich aber in vielen Befunden und Studien der vorliegenden Arbeit ein hoher Grad prosozialen Verhaltens auch in High-Cost-Situationen: in der Präferenz einer solidarischen gesetzlichen Krankenversicherung insbesondere bei Wählern von Bündnis 90/Die Grünen (Sekundäranalyse in Studie 4); bei Aufschlüsselung nach finanziellen Profiteuren und Schlechtergestellten durch pauschale Krankenkassenbeiträge (Sekundäranalyse Pappi & Shikano [2004] in Studie 4); im *Identifiable-Victim-Effekt* in Form der Entscheidung vieler Bessergestellter, erhebliche finanzielle Einbußen in Kauf zu nehmen (Voting-Studie 3); sowie im Öffentliches-Gut-Spiel (Studie 1) im *No-Feedback Treatment*, wenn Vpn kumuliert über 20 Runden erhebliche Beiträge (ohne Sanktionsgefahr) an andere Mitspieler verteilen. Variiert die Auszahlungshöhe jedoch innerhalb des gleichen Experimentaldesigns, zeigt sich beispielsweise in Voting-Studie 3 eigennutzmaximierendes Verhalten durchschnittlich umso stärker, je mehr Geld faires Verhalten kostet, und auch in Studie 5 (Krankenversicherungsbeiträge) ist eine positive Korrelation zwischen monetärem Nutzen und Verhalten nachweisbar. In Studie 1 ist ebenfalls eine Tendenz zu Eigennutz feststellbar, je höher die akkumulierten Verdienste aller Runden steigen. Die RCT erweist sich somit umso leistungsfähiger, je höher die Kosten sind. Dieser Effekt verstärkt sich in eindeutigen und in ökonomisch geprägten Situationen. Demgegenüber erweisen sich Rahmentheorien und Konzepte mittlerer und kleiner Reichweite in Low-Cost-Situationen als besonders erklärungsstark.

Granularität

In vier Studien erweisen sich Eigennutz auf Aggregatebene als robuster Hauptfaktor und soziale Präferenzen als Nebeneffekt. Je häufiger ein Experiment durchgeführt wird (Studie 1), je stärker die Ergebnisse eines Treatments über Gruppen (Studie 1), Treatments (Studien 1, 3, 4 und 5) oder die Zeit (Spendenoption in Studie 3) aggregiert werden, desto stärker zeigt sich die RCT. Je kleiner jedoch die Granularität der Untersuchungsobjekte (Studie 2), der Datenanalyse (singuläres Verhalten und Einstellungen in den Studien 1, 3 und 5) und deren Gesamtschau auf Individualebene (Studien 1, 3 und 5) sind, desto leistungsfähiger erscheinen Rahmentheorien und Konzepte mittlerer und kleiner Reichweite (z. B. *Identifiable-Victim-Effekt* in der Voting-Studie 3, Klassifikation der Fixationsdauern des Augenbewegungsverhaltens in Studie 2). Werden die Vpn umfassend auf der Mikroebene betrachtet, zeigen sich statt stringenter Verhaltens- und Begründungsstrategien heterogene, dynamische mentale Prozesse und Verhaltensmuster. In der Voting-Studie 3 wird dieser paradox erscheinende Zusammenhang plausibel, da die starke Benachteiligung (negative Geldauszahlung) einer sehr kleinen Subgruppe einen starken *Identifiable-Victim-Effekt* erzeugt, aufgrund dessen viele Vpn gegen die simultan eigennutzmaximierende,

relativ gleichmäßige und gruppeneffiziente Verteilung stimmen. Dennoch wird die Koalition aus finanziell eigennützigem Abstimmungsverhalten der benachteiligten Spieler und finanziell uneigennützigem *Identifiable-Victim-Effekt*-Spielern aufgrund der Gruppenzusammensetzung (nur wenige stimmberechtigte Benachteiligte) und des hohen Anteils an Spielern, die in High-Cost-Situationen eigennutzmaximierend handeln, überstimmt. Es kommt daher auf Aggregatebene zu dem von der RCT prognostizierten Ergebnis. Jedoch entsprechen viele Ergebnisse auf Individualebene nicht der RCT. Beispielsweise verteilen nach Abschluss des Experimentes in der Umverteilungsoption einige Vpn, die vorher für die eigennutzmaximierende Option gestimmt hatten, einen Teil ihres Geldes an die Subgruppe der Benachteiligten.

Korrespondenz von Verhalten, Präferenzen und Motiven

Die Ambivalenz der Leistungsfähigkeit der RCT zeigt sich auch in der Diskrepanz zwischen Verhalten, Präferenzen und Motiven. Je stärker eine Einstellung auf ein konkretes Verhalten und dessen faktische Konsequenzen (z. B. mehr Geld) gerichtet ist, desto leistungsfähiger erscheint die RCT. Hinter fairem Verhalten verbergen sich eigennützige Motive, beispielsweise in Studie 3, wenn die Vpn für eine gleichverteilte Option stimmen, da sie sich in der benachteiligten Subgruppe befinden. Vpn aus der Subgruppe der Benachteiligten geben dabei an, die gleichverteilte (faire) Option nicht aus Fairness, sondern aus Eigennutz gewählt zu haben. Im Strategiemodus haben jedoch teilweise dieselben Vpn angegeben, dass sie in der Rolle einer begünstigten Person die gleichverteilte Option aus Fairness und nicht aus Eigennutz gewählt hatten. Intrapersonell wurde somit von der gleichen Person die gleichverteilte Option aus zwei unterschiedlichen Motiven gewählt. Umgekehrt begründen viele Vpn ihr prosoziales Verhalten mit Eigennutz, auch wenn dieser im Verhalten nicht erkennbar ist.

Bezüglich intrapersonaler Konsistenz des Entscheidungsverhaltens und situativer Persönlichkeitszüge sind sowohl im Treatment *Inverses Diktatorspiel* (Studie 5) als auch im Voting-Experiment (Studie 3) individuelle Verhaltensmuster zu beobachten, die der RCT widersprechen. Spieler aktivieren im jeweiligen Szenario bzw. in der jeweiligen Personenrolle unterschiedliche situative Persönlichkeitszüge, Entscheidungstypen und Bewertungskriterien. Dies zeigt sich auch in den geäußerten Entscheidungsgründen. Im Voting-Experiment korrelieren die Verhaltensgründe (z. B. Fairness) teilweise nicht mit dem Verhalten (z. B. Wahl einer eigennützigen Option). Die geringe Korrelation zwischen grundsätzlicher Einstellung zu den Verhaltensoptionen, individuell angegebener Begründung für das Verhalten und tatsächlich erfolgtem Entscheidungsverhalten wird in der Literatur zur Methodik der empirischen Umfrageforschung (Ajzen & Sexton, 1999; Schmitt, 1990) damit erklärt, dass Individuen versuchen, sozial erwünschtes, konsistentes Antwortverhalten zu zeigen (Esser, 1986), oder auf den *ökonomischen Reflex* zurückfallen. Diese Verzerrung des Antwortverhaltens erklärt jedoch nur einen Teil der Befunde. Wenn Individuen fair handeln, jedoch nicht-soziale Motive oder theorieferne Gründe angeben,

kann dies auch aus der Heterogenität mentaler Modelle resultieren, die nicht den bisher diskutierten, üblichen Erklärungsschemata zuordenbar sind. So ist beispielsweise der Entscheidungsgrund *Orientierung an den Anderen* zwar theoriefern, da er nicht auf konditionale Kooperation, sondern wertrational und habituell auf ein an der Gruppenmehrheit und deren Normen orientiertes Verhalten abzielt, aber zur Outcome-Maximierung im experimentellen wiederholten Öffentliches-Gut-Spiel (Studie 1) noch in gewisser Weise nachvollziehbar und sinnvoll. Dieser Verhaltensgrund ist jedoch in einem one-shot-Spiel, also einem Spiel über nur eine Runde wie in Studie 3, durch keine RCT-nahe Theorie erklärbar. Die Bandbreite individueller Verhaltensmuster ist hierbei teilweise auch durch die Existenz situativer Persönlichkeitszüge und die relativ lockere Verbindung zwischen mentalen Prozessen und Verhalten erklärbar.

Einen Sonderfall stellen inkonsistente Einstellungs- und Verhaltensmuster dar, die in der RCT nicht zulässig sind. Wenn Individuen aus Eigennutz die faire Option wählen und aus Fairness die eigennützige Option, kann dies an der subjektiven Interpretation von Situationen und Handlungsmotiven liegen. Jedoch geben in der Studie 3 viele Vpn an, eine Option sowohl aus Fairness als auch aus Eigennutz gewählt zu haben. Aus sozialen Präferenzen finanzielle Nachteile in Kauf zu nehmen und diese Entscheidung als fair zu bewerten, widerspricht dem simultan genannten Grund, damit den Eigennutz maximieren zu wollen. In Studie 3 (Voting) wird nach Abschluss des Experimentes die Möglichkeit gegeben, einen beliebigen Anteil des im Experiment verdienten Geldes an die anderen Mitspieler oder als externe Spende an die Hilfsorganisation „Ärzte ohne Grenzen“ umzuverteilen. Wird (wie von der RCT axiomatisiert) Konsistenz angenommen und das Verhalten während des Experimentes zur Typisierung von Spielern als eigennützig oder fairnessorientiert herangezogen, ist nicht erklärbar, warum kein Zusammenhang zwischen dieser Typisierung und der Reallokation besteht. Insbesondere haben die Wähler der solidarischen Option kaum Geld abgegeben, wenn sie sich später in der Gruppe der Bessergestellten wiedergefunden haben. Umgekehrt waren Wähler der eigennützigen Option bereit, einen Teil ihres Einkommens umzuverteilen.

Eindeutigkeit der Situation

Die RCT erweist sich in Situationen mit eindeutiger Reizvorlage als erklärungsstark, demgegenüber entfalten Rahmentheorien in uneindeutigen Situationen eine hohe Prädiktorkraft. In Studie 1 bestehen im Treatment ohne Feedback keine starken äußeren Hinweisreize, so dass sich ein Drittel der Spieler an inneren Überzeugungen orientiert, die nur einen unscharfen Bezug zur RCT haben, zwei Drittel der Vpn jedoch scheinbar orientierungslos unklaren Heuristiken und Mustern folgen, die gegen die RCT sowie die hieraus abgeleiteten Typen (z. B. *konditionaler Kooperator*) verstoßen. Als verbindende Elemente erweisen sich der Grundmechanismus des *Match* und die Variabilität mentaler Modelle, beispielsweise in Form von Episoden als Bewusstseinsselemente, die Reize, Reaktionen und Konsequenzen umfassen.

Dagegen führen eindeutige Situationen (z. B. experimentelle Auktionen) und ein klares Framing (z. B. die Betonung des ökonomischen Kontextes wie im Wall Street Game [Liberman, Samuels, & Ross, 2004]) zu RCT-kompatiblen Verhalten. Um der Forschungsfrage dieser Dissertation gerecht zu werden und die Leistungsfähigkeit unterschiedlicher handlungstheoretischer Forschungstraditionen untersuchen zu können, wurden bewusst keine eindeutigen, RCT-nahen Situationen und Designs gewählt.

Grad der Erweiterung der RCT

Je stärker die RCT mit ihrem Rationalitäts- und Nutzenbegriff erweitert wird, desto mehr Ergebnisse können eingeschlossen werden: Altruistische Präferenzen (Studie 1), wertrationales Handeln (Studien 1, 3 und 4), Orientierung an der Gruppeneffizienz (Studien 1 und 3), Orientierung an der Salienz (Präferenz für eine solidarische Finanzierung der gesetzlichen Krankenversicherung in Studie 4) beziehungsweise an der mangelnden Salienz (z. B. Marginalität des eigenen Beitrags in Studie 1). Wenn die RCT immer mehr erweitert wird, befinden sich kaum noch Befunde außerhalb dieser Theorie. Jedoch sinkt damit gleichzeitig die Erklärungskraft, da zunehmend *ex post* begründet werden muss, warum ein Individuum in einer spezifischen Situation altruistisch, ungleichheitsavers, eigennützig, antisozial oder willkürlich handelt.

Rahmentheorien

Während Studie 4 zur Krankenkassenfinanzierung Rahmentheorien nur indirekt stützt, stärkt die Augenbewegungsstudie (Studie 2) die Annahme der Unabhängigkeit von Entscheidungs- und Verhaltenssystem. Sowohl im Voting-Experiment (Studie 3) als auch im Diktatorspiel (Studie 5) werden sowohl die Bedeutung heterogener Entscheidungsheuristiken als auch deren Zusammenwirken deutlich. Die Entscheidungsheuristiken orientieren sich nicht auf einer eindimensionalen, mit der RCT kompatiblen Bewertungsachse, die intertemporal und intersituativ stabil und durch die monetären Auszahlungen determiniert ist. Im Widerspruch zu Theorien der Ungleichheitsaversion (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) ist es in der Voting-Studie 3 in einigen Treatments hoch relevant, wie die Auszahlungen zwischen den Mitspielern verteilt werden. Im Diktatorspiel (Studie 5) wie auch in Studie 3 wird deutlich, dass Rahmentheorien die Ergebnisse zwar erklären können, jedoch einer inhaltlichen Füllung bedürfen, um zu zeigen, welche konkreten mentalen Modelle, Episoden und Gerechtigkeitsprinzipien von Individuen aktiviert werden. Studie 1 zum Öffentliches-Gut-Spiel zeigt die Leistungsfähigkeit von Rahmentheorien im Anwendungsfall limitierter Informationen und Hinweisreize bezüglich des Verhaltens der Mitspieler. Individuen greifen auf eine Vielfalt an *matchenden* Verhaltensmustern, Heuristiken und Episoden zurück, die sich aufgrund innerer Prozesse im Zeitverlauf ändern können (situative Persönlichkeitszüge und Reframing). Die besondere Leistungsfähigkeit von Rahmentheorien entfaltet sich bei der Betrachtung und

Analyse auf kleiner Granularität. Dem Verhalten widersprechende, untereinander widersprüchliche Verhaltensgründe werden durch Rahmentheorien und neurowissenschaftliche Ansätze plausibel, da in diesen eine Vielzahl von Motiven und Heuristiken subjektiv angemessen sein kann. Das Konzept mentaler Modelle und des *Match* erscheinen geeignet, eine große Bandbreite an Verhalten zu modellieren. Die von der RCT angenommene Existenz einer zentralen Exekutive ist hierfür nicht zwingend erforderlich.

Methode

Vier Studien basieren auf spieltheoretischen Experimenten oder Experimentserien, für die im Anschluss die Entscheidungsgründe der Teilnehmer erhoben wurden. Dieses Vorgehen wurde gewählt, da eine empirische Erhebung von Präferenzen und Motiven aus der methodischen Perspektive der empirischen Sozialforschung notwendig erscheint. Aus der Sicht von Rahmentheorien stellen die Spielsituation und die Persönlichkeit des Spielers keine exogen gegebenen Parameter dar, sondern sind das Ergebnis teils aktiver Auswahl- und Aktivierungsprozesse. Die Methode der vorliegenden Arbeit umfasst des Weiteren Sekundäranalysen von Experimenten und Umfragen sowie die Messung von Augenbewegungen. In Studie 1 werden zudem zwei methodische Probleme im Öffentliches-Gut-Spiel von Fehr und Gächter (2000) diskutiert und anschließend empirisch untersucht: das strukturell verzögerte Feedback und Gruppen als eigenständige Treatments, die in statistischen Analysen nur dann aggregierbar sind, wenn sich die jeweiligen Feedbackverläufe nicht grundlegend unterscheiden. In der Voting-Studie 3 besteht für Teilnehmer nach Abschluss des Experimentes die Möglichkeit, den erspielten Geldbetrag an die Mitspieler umzuverteilen oder an eine Hilfsorganisation zu spenden. Durch diese Methode sowie die ergänzende Erhebung der Verhaltensgründe werden im Verhalten intraindividuelle Inkonsistenzen beobachtbar. Studie 5 (Diktatorspiel) zeigt mit der Strategiemethode (Selten, 1967; Mitzkewitz & Nagel, 1993) intraindividuelle Entscheidungsmuster. Anhand der Strategiemethode lassen sich bereits an einzelnen Individuen Abweichungen von den Prognosen der RCT beobachten. Zudem hat es sich für die Entwicklung des Experimentaldesigns der Studien als fruchtbar erwiesen, anhand von Rahmentheorien und Konzepten kleiner Reichweite gezielt inkongruentes Verhalten, einen Konflikt unterschiedlicher Handlungsmuster oder einen dynamischen Wechsel im Zeitverlauf (Reframing) zu identifizieren.

Die vorliegende Evidenz stützt die These, dass Individuen aktiv ihre Wahrnehmung einer Situation und die angemessenen situativen Persönlichkeitszüge wählen. Daraus folgt als gemeinsamer Ertrag der Studien die Empfehlung, Verhaltensmessungen immer um eine Erhebung der mentalen Prozesse zu ergänzen. Konkret: Da Individuen eine Spielsituation unterschiedlich wahrnehmen und unterschiedliche Motive zum gleichen Verhalten führen können, kann aus Outcome-Messungen (Verhalten) nicht zwangsläufig auf mentale Prozesse (Motive) rückgeschlossen werden.

Limitationen und Fazit

Menschliches Verhalten wird in den fünf Studien der vorliegenden Dissertation in verschiedenen Anwendungsfeldern untersucht. Die Studien umfassen politische Einstellungen bezüglich pauschaler versus einkommensabhängiger Finanzierungsoptionen im Gesundheitswesen, spieltheoretische Labor- und Hörsaalexperimente (Öffentliches-Gut-Spiel, Voting-Experimente in großen und kleinen Gruppen, Diktatorspiel) bis hin zu Fixationsdauern des Augenbewegungsverhaltens. Um die übergeordnete Forschungsfrage zu beantworten, wie menschliches Verhalten gesteuert wird, werden insbesondere in den vier Studien, die sich mit spieltheoretischen Experimenten beschäftigen, die Prognosen der RCT und ihrer Erweiterungen den Prognosen von alternativen Mechanismen der Handlungssteuerung gegenübergestellt. In allen Studien der Dissertation zeigt sich teils massive empirische Evidenz, die nicht im Einklang mit der RCT steht. Die Abweichung von der RCT ist hierbei nicht zufällig, sondern systematisch beobachtbar. Je umfassender und kleingranularer Individuen im sozialen Feld betrachtet werden, je uneindeutiger und vom ökonomischen Kontext ferner eine Situation ist, je geringer die individuellen (finanziellen) Kosten sind, desto eher widersprechen die Befunde den Prognosen der RCT. In diesen Bereichen zeigen Rahmentheorien ihre große Leistungsfähigkeit. Das Heranziehen von Rahmentheorien als allgemeiner Erklärungsrahmen erweist sich insbesondere dort als fruchtbar, wo die RCT eine relativ geringe Prognosekraft besitzt. Der Mechanismus des *Match* des MFS erweist sich in diesem Kontext als leistungsfähiger im Vergleich zum Nutzenmaximierungsprinzip der RCT. Andererseits zeigt sich insbesondere im ökonomischen Kontext und auf Aggregatebene die große Leistungsfähigkeit der RCT. Die Leistungsfähigkeit resultiert nicht nur aus primären, sondern auch aus sekundären Effekten, da beispielsweise rationales Verhalten eine kulturelle Norm darstellt. Aus der Perspektive von Rahmentheorien stellt die RCT einen Spezialfall dar, bei dem Verhaltensskripte und Episoden nicht automatisch abgerufen werden, sondern es zu teils extensiven Abwägungsprozessen kommt.

Allerdings sind für eine Übertragung der Ergebnisse der Dissertation auf andere Untersuchungsgegenstände und Anwendungsfelder die nachfolgend skizzierten Limitationen zu beachten. Während die Studien 2 und 4 Sekundäranalysen darstellen, die sich auf große Fallzahlen stützen, ergeben sich in den Studien 1, 3 und 5 aufgrund der geringen Fallzahl Limitationen bezüglich der internen Validität und Reliabilität. Studie 1 stützt das Argument schlechter Replizierbarkeit, da in der Replikationsstudie des Öffentlichen-Gut-Spiels von Fehr und Gächter (2000) ein völlig anderes Beitragsniveau nachgewiesen wird. In der ersten Runde unterscheiden sich die beiden Untersuchungen zunächst nicht signifikant voneinander, sodass hieraus der Umkehrschluss gezogen werden kann, dass in der Literatur möglicherweise generell sehr starke Limitationen und methodische Fehler bestehen, die bisher unzureichend beachtet wurden. Da zudem die (explorative) Erforschung individuellen Verhaltens der zentrale Untersuchungsgegenstand der Dissertation ist, wird als Forschungsstrategie eine Kombination der experimentellen Methode mit quantitativen und qualitativen Methoden der empirischen Sozialforschung gewählt,

um Individuen möglichst umfassend zu betrachten und eine möglichst hohe Validität der empirischen Befunde zu erreichen. Durch die Verhaltensmessung im Stratiemodus, Experimentserien mit den gleichen Versuchspersonen und die Erhebung der Entscheidungsgründe wird durch diesen Mixed-Methods-Ansatz ein tiefer Einblick möglich. Daher stellen die vorliegenden Studien trotz ihrer Limitationen einen Ausgangspunkt für zukünftige Serien experimenteller Untersuchungen dar.

Weitere Limitationen sind die in Studie 4 aus der repräsentativen Bevölkerungsstudie gezogenen Schlussfolgerungen, da eine Vielzahl von Drittvariablen nicht kontrolliert wurde, insbesondere der Zusammenhang zwischen Einstellungen und Verhalten. Aus dem sehr heterogenen Gesamtbild resultiert eine eingeschränkte Übertragbarkeit und Anwendbarkeit der Untersuchungsergebnisse auf konkrete, reale Themen und Sachverhalte. Dies kann als Limitation gesehen werden oder auch als grundlegende Aussage, dass die Übertragung von experimentellen Untersuchungsergebnissen auf reale Sachverhalte mit Vorsicht zu betrachten ist, da es ex ante nahezu unmöglich erscheint, beispielsweise für den *Identifiable-Victim-Effekt* die Randbedingungen und situativ korrekte Schwellenwerte zu prognostizieren.

Durch die methodische und theoretische Breite der Dissertation, aber auch durch die Einbeziehung und Nachnutzung von Forschungsprimärdaten einer breiten und soliden Datenbasis, ergibt sich in der Gesamtschau ein robustes Gesamtbild mit der Folgerung, dass die RCT als Spezialfall von Rahmentheorien pragmatisch und anwendungsfallspezifisch in den Bereichen genutzt werden sollte, in denen sie eine besonders hohe Prädiktorkraft aufweist (z. B. im ökonomischen Kontext oder zur Erklärung von Befunden auf Aggregatebene). Demgegenüber sollten Rahmentheorien stärker zur Erklärung individuellen sowie intra-individuellen Entscheidungsverhaltens insbesondere in uneindeutigen und Low-Cost-Situationen genutzt werden. Hier sind jedoch noch erhebliche Herausforderungen zu überwinden, da Rahmentheorien durch Konzepte und Heuristiken mittlerer und kleiner Reichweite inhaltlich gefüllt werden müssen. Auf Grundlage der teils explorativen Ergebnisse der vorliegenden Dissertation erscheint es lohnenswert zu untersuchen, inwiefern sich Rahmentheorien auch in Experimentserien und Metaanalysen mit großen Fallzahlen als erklärungsstark erweisen. Das möglicherweise wichtigste Ergebnis der Dissertation könnte jedoch die Erkenntnis sein, dass sich die Verhaltensforschung trotz extensiver Erforschung in vielen Fachdisziplinen und anhand unterschiedlicher Methoden nicht etwa einem Ende nähert, sondern erst an ihrem Anfang zu stehen scheint, wenn es darum geht, die unüberschaubare Fülle immer neuer Aspekte menschlichen Verhaltens und der begleitenden mentalen Prozesse deutend zu erklären, beispielsweise die synthetische (statt analoge) Integration von Informationen. „Intellectus igitur, qui non est veritas, numquam veritatem adeo praecise comprehendit, quin per infinitum praecisius comprehendere possit, habens se ad veritatem sicut polygonia ad circulum, quae quanto inscripta plurium angulorum

fuerit, tanto similior circulo [; ...] et quanto in hac ignorantia profundius docti fuerimus, tanto magis ipsam accedimus veritatem.“²

Literatur

- Abbink, K., & Sadrieh, A. (2009). The pleasure of being nasty. *Economics Letters*, 105(3), 306-308. doi: 10.1016/j.econlet.2009.08.024
- Ajzen, I., & Sexton, J. (1999). Depth of processing, belief congruence, and attitude-behavior correspondence. In S. Chaiken & Y. Trope (Hrsg.), *Dual-process theories in social psychology* (S. 117-138). New York, NY: Guilford Press.
- Bell, S. (2011). Do we really need a new 'constructivist institutionalism' to explain institutional change? *British journal of political science*, 41(4), 883-906.
- Best, H., & Kroneberg, C. (2012). Die Low-Cost-Hypothese. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 64(3), 535-561. doi:10.1007/s11577-012-0174-5
- Betz, D. (2020). *Konditionale Kooperation in experimentellen Öffentliches-Gut-Spielen mit und ohne Feedback: Erklärung oder Artefakt?* Manuskript eingereicht zur Publikation.
- Betz, D. Oestreich, M., & Biniossek, C. (2020). *The Squeaky Wheel Gets the Grease – Reference Point Variations in Game Theoretic Experimental Dictator Games*. Manuskript eingereicht zur Publikation.
- Biniossek, C., & Betz, D. (2018). *Der Einfluss unterschiedlicher Gerechtigkeitsprinzipien auf Mehrheitsentscheidungen in einem spieltheoretischen Social-Good-Experiment*. (GESIS Papers, 2018/20). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. doi: 10.21241/ss0ar.58261
- Biniossek, C., Betz, D., Fetschenhauer, D., & Lungen, M. (2007). Fairness ohne Eigennutz. Handlungsannahmen in der ökonomischen Theorie und Auswirkungen auf Reformoptionen im Gesundheitswesen. *Sozialer Fortschritt*, 56(9-10), 252-258.
- Blatter, J. K., Janning, F., & Wagemann, C. (2007). *Qualitative Politikanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/978-3-531-90716-1
- Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *The American Economic Review*, 90(1), 166-193.
- Brandstätter, H. & Güth, W. (1994). Introduction to essays on economic psychology. In H. Brandstätter & W. Güth (Hrsg.), *Essays on Economic Psychology* (S. 1-29). Berlin/Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-48621-0
- Braun, D. (2013). *Theorien rationalen Handelns in der Politikwissenschaft: Eine kritische Einführung* (Band 25). Wiesbaden: Springer-Verlag.
- Braun, N., & Wolbring, T. (2012). Ende der Ökonomisierung? *Soziale Welt*, 379-400.

2 „[Der Geist] verhält sich zur Wahrheit wie ein dem Kreis eingeschriebenes Vieleck, welches dem Kreis um so ähnlicher sein wird, je mehr Winkel es hat [; ...] und je tiefer wir in dieser Unwissenheit wissend und weise geworden sein werden, um so mehr werden wir uns der Wahrheit selbst nähern.“ (Nikolaus von Kues in der Übersetzung von Dupré, https://urts99.uni-trier.de/cusanus/content/fw.php?werk=13&ln=dupre&dupre_fw=1 [Zugegriffen: 07. Juni 2020]).

- Brosig-Koch, J., Riechmann, T., & Weimann, J. (2007). *Selfish in the end? An investigation of consistency and stability of individual behavior* (FEMM Working Papers Nr. 05). Magdeburg: Otto-von-Guericke Universität.
- Buchanan, J. M., & Tullock, G. (1962). *The calculus of consent* (Band 3). Ann Arbor, MI: University of Michigan press. doi: 10.3998/mpub.7687
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology*, 39(5), 752-766. doi: 10.1037/0022-3514.39.5.752
- Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Hrsg.), *Unintended thought* (S. 212-252). New York, NY: Guilford Press.
- Christoph, B., & Ullrich, C. G. (2006). Die GKV in den Augen der Bürger: Wahrnehmung des Lastenausgleichs und Bewertungen von Reformoptionen. *Sozialer Fortschritt*, 75-83.
- Coleman, James S. (1986): Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology* 91, 1309-1335.
- Dahrendorf, R. (1958). *Homo Sociologicus. Ein Versuch zur Geschichte, Bedeutung und Kritik der Kategorie der sozialen Rolle*. Opladen: Westdeutscher Verlag.
- Diekmann, A., & Preisendörfer, P. (1992). Persönliches Umweltverhalten: Diskrepanzen zwischen Anspruch und Wirklichkeit. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 44(2), 226-251.
- Dowding, K., & King, D. (Hrsg.) (1995). *Preferences, institutions, and rational choice*. Oxford: Clarendon Press.
- Downs, A. (1957). *An Economic Theory of Democracy*. New York, NY: Harper.
- Easton, D. (1965). *A systems analysis of political life*. New York: Wiley.
- Edwards, W. (1954). The theory of decision making. *Psychological bulletin*, 51(4), 380-417. doi: 10.1037/h0053870
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4), 583-610. doi: 10.1007/s10683-011-9283-7
- Erb, H. P., Kruglanski, A., Chun, W. Y., Pierro, A., Mannetti, L., & Spiegel, S. (2003). Searching for commonalities in human judgement: The parametric unimodel and its dual mode alternatives. *European review of social psychology*, 14(1), 1-47. doi: 10.1080/10463280340000009
- Esser, H. (1986). Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. (ZUMA-Arbeitsbericht, 1986/02). Mannheim: Zentrum für Umfragen, Methoden und Analysen – ZUMA.
- Esser, H. (1993). *Soziologie. Allgemeine Grundlagen*. Frankfurt am Main/New York: Campus.
- Esser, H. (1996). Die Definition der Situation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 48(1), 1-34.
- Esser, H. (1999). *Soziologie: Spezielle Grundlagen. Situationslogik und Handeln* (Band 1). Frankfurt a. M./ New York: Campus.

- Esser, H. (2001a). Soziologie. Spezielle Grundlagen, Band 1: Situationslogik und Handeln. Frankfurt a. M./New York: Campus.
- Esser, H. (2001b). Soziologie. Spezielle Grundlagen. Band 6: Sinn und Kultur. Frankfurt a. M./ New York: Campus.
- Esser, H. (2002a). In guten wie in schlechten Tagen? Das Framing der Ehe und das Risiko zur Scheidung. Eine Anwendung und ein Test des Modells der Frame-Selektion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 54, 27-63.
- Esser, H. (2002b). Ehekrisen: Das (Re-)Framing der Ehe und der Anstieg der Scheidungsraten. *Zeitschrift für Soziologie*, 31(6), 472-496.
- Esser, H. (2003). Der Sinn der Modelle. Antwort auf Götz Rohwer. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 55, 359. doi:10.1007/s11577-003-0055-z
- Esser, H. (2010a). Das Modell der Frame-Selektion. Eine allgemeine Handlungstheorie für die Sozialwissenschaften. Soziologische Theorie kontrovers. *Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 50, 45-61.
- Esser, H. (2010b). Sinn, Kultur, Verstehen und das Modell der soziologischen Erklärung. In M. Wohlrab-Sahr (Hrsg.), *Kulturosoziologie* (S. 309-335). VS Verlag für Sozialwissenschaften.
- Esser, H. (2018). Sanktionen, Reziprozität und die symbolische Konstruktion einer Kooperations-„Gemeinschaft“. *Zeitschrift für Soziologie*, 47(1), 8-28. doi: 10.1515/zfsoz-2018-1001
- Esser, H. (2019). Rational Choice oder Framing?. In J. Mayerl, T. Krause, A. Wahl, M. Wuketich (Hrsg.), *Einstellungen und Verhalten in der empirischen Sozialforschung*. Wiesbaden: Springer VS. doi: 10.1007/978-3-658-16348-8_3
- Esser, H., & Kroneberg, C. (2015). An integrative theory of action: the model of frame selection. In E. J. Lawler, S.R. Thye, & J. Yoon (Hrsg.), *Order at the Edge of Chaos: Social Psychology and the Problem of Social Order* (S. 63-85). New York, NY: Cambridge University Press.
- Fallucchi, F., Luccasen, R. A., & Turocy, T. L. (2019). Identifying discrete behavioural types: a re-analysis of public goods game contributions by hierarchical clustering. *Journal of the Economic Science Association*, 5(2), 238-254. doi: 10.1007/s40881-018-0060-7
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Hrsg.), *Advances in experimental social psychology*, 23(75-109). London, GB; San Diego, CA: Academic Press.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980-994. doi: 10.1257/aer.90.4.980
- Fehr, E., & Schmidt, K. M. (1999). A Theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114(3): 817-868. doi: 10.1162/003355399556151
- Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism – experimental evidence and new theories. In S. C. Kolm & J. M. Ythier (Hrsg.), *Handbook of the economics of giving, altruism and reciprocity*, (Band 1, S. 615-691). Elsevier Science. doi: 10.1016/S1574-0714(06)01008-6
- Fishbein, M., & Ajzen, I. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological review*, 81(1), 59-74. doi: 10.1037/h0035872

- Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.
- Fochmann, M., Jahnke, B., & Wagener, A. (2019). Does the reliability of institutions affect public good contributions? Evidence from a laboratory experiment. *Scottish Journal of Political Economy*, 66(3), 434-458. doi: 10.1111/sjpe.12197
- Fochmann, M., & Wolf, N. (2019). Framing and salience effects in tax evasion decisions – An experiment on underreporting and overdeducting. *Journal of Economic Psychology*, 72, 260-277. doi: 10.1016/j.joep.2019.03.005
- Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behavior*, 6(3), 347-369. doi: 10.1006/game.1994.1021
- Frings, C., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Hommel, B., Kiesel, A., Kunde, W., Mayr, S., Moeller, B., Möller, M., Pfister, R., & Philipp, A. (2020a). Merkmalsintegration und Abruf als wichtige Prozesse der Handlungssteuerung – eine Paradigmen-übergreifende Perspektive. *Psychologische Rundschau*, 71, 1-14. doi: 10.1026/0033-3042/a000423
- Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Kiesel, A., Kunde, W., Mayr, S., Moeller, B., Möller, M., Pfister, R., & Philipp, A. (2020b). Binding and Retrieval in Action Control (BRAC). *Trends in Cognitive Sciences*, 1-13. doi: 10.1016/j.tics.2020.02.004.
- Frith, C. D., & Haggard, P. (2018). Volition and the brain – revisiting a classic experimental study. *Trends in Neurosciences*, 41(7), 405-407. doi: 10.1016/j.tins.2018.04.009
- Galley, N., Betz, D., & Biniossek, C. (2015). Fixation durations: Why are they so highly variable. In T. Heinen (Hrsg.), *Advances in Visual Perception Research* (S. 83-106). New York, NY: Nova Science.
- Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.
- Goerg, S. J., Rand, D., & Walkowitz, G. (2019). Framing effects in the prisoner's dilemma but not in the dictator game. *Journal of the Economic Science Association*, 1-12. doi: 10.1007/s40881-019-00081-1
- Habermas, J. (1981). *Theorie des kommunikativen Handelns*. Frankfurt: Suhrkamp.
- Hagauer, H., Kittel, B., & Schwaninger, M. (2019). Fairness in Bargaining: How Self-selected Frames Affect the Fairness of Negotiation Outcomes. In M. Debus, T. Tepe, & J. Sauermann (Hrsg.) *Jahrbuch für Handlungs- und Entscheidungstheorie* (S. 3-31). Wiesbaden: Springer VS.
- Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, 126(1), 128-133.
- Jahn, D. (2013). *Einführung in die vergleichende Politikwissenschaft*. Wiesbaden: Springer VS.
- Jenni, K. E., & Loewenstein, G. (1997). Explaining the “Identifiable Victim Effect”. *Journal of Risk and Uncertainty*, 14(3), 235-257.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.

- Kaiser, A. (1997). Types of democracy: From classical to new institutionalism. *Journal of Theoretical Politics*, 9(4), 419-444.
- Kaiser A. (2001) Die politische Theorie des Neo-Institutionalismus: James March und Johan Olsen. In A. Brodocz & G. S. Schaal (Hrsg.), *Politische Theorien der Gegenwart II. Eine Einführung*. (S. 253-282). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/978-3-663-12320-0_9
- Kittel, B., Luhan, W., & Morton, R. (Hrsg.) (2012). *Experimental political science: Principles and practices*. Hampshire, GB: Palgrave Macmillan.
- Klein, M., & Falter, J. W. (2003). *Der lange Weg der GRÜNEN: eine Partei zwischen Protest und Regierung*. München: Beck.
- Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. In A. Kurtz (Hrsg.), *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 284(1), S. 1-17). Springer. doi: 10.1007/BF00412364
- Kroneberg, C. (2005). Die Definition der Situation und die variable Rationalität der Akteure. Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie*, 34(5), 344-363.
- Kroneberg, C. (2011). *Die Erklärung sozialen Handelns: Grundlagen und Anwendung einer integrativen Theorie*. Wiesbaden: Springer.
- Kroneberg, C. (2014). Frames, scripts, and variable rationality: An integrative theory of action. In G. Manzo (Hrsg.), *Analytical Sociology. Actions and Networks* (S. 97-123). Chichester, GB: Wiley.
- Kruglanski, A. W., Pierro, A., Mannetti, L., Erb, H. P., & Chun, W. Y. (2007). On the parameters of human judgment. *Advances in experimental social psychology*, 39, 255-303. doi: 10.1016/S0065-2601(06)39005-3
- Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and social psychology bulletin*, 30(9), 1175-1185.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, 106(3), 623-642. doi: 10.1093/brain/106.3.623
- Lijphart, A. (1971). Comparative politics and the comparative method. *American political science review*, 65(3), 682-693. doi: 10.2307/1955513
- Lijphart, A. (1975). II. The comparable-cases strategy in comparative research. *Comparative political studies*, 8(2), 158-177. doi: 10.1177/001041407500800203
- Manzo, G. (2013). Is rational choice theory still a rational choice of theory? A response to Opp. *Social Science Information*, 52(3), 361-382. doi: 10.1177/0539018413488477
- March, J. G., & Olsen, J. P. (1983). The new institutionalism: Organizational factors in political life. *American political science review*, 78(3), 734-749. doi: 10.2307/1961840
- March, J. G., & Olsen, J. P. (2006). Elaborating the "new institutionalism". In R. A. W. Rhodes, S. A. Binder, & B. A. Rockman (Hrsg.), *The Oxford handbook of political institutions*, (5, S. 3-20). Oxford, GB: Oxford University Press.
- Mill, J. S. (1843). A system of logic, ratiocinative and inductive, being a connected view of the principles of evidence, and the methods of scientific investigation (Band 1). London:

- John W. Parker. Abgerufen von https://archive.org/details/bub_gb_y4MEAAAQAAJ/page/n1/mode/2up
- Mitzkewitz, M. & Nagel, R. (1993). Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory*, 22(2), 171-198.
- Morton, R. B., & Williams, K. C. (2010). *Experimental political science and the study of causality: From nature to the lab*. Cambridge, MA: Cambridge University Press.
- Musgrave, R. A. (1959). *Theory of public finance: a study in public economy*. New York, NY: McGraw-Hill.
- Nann, M., Cohen, L. G., Deecke, L., & Soekadar, S. R. (2019). To jump or not to jump – The Bereitschaftspotential required to jump into 192-meter abyss. *Scientific reports*, 9(1), 1-9. doi: 10.1038/s41598-018-38447-w
- Niskanen, W. A. (1971). *Bureaucracy and Representative Government*. Chicago, IL: Aldine-Atherton
- Olson, M. (1965). *The logic of collective action. Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.
- Opp, K. D. (1978). Das „ökonomische Programm“ in der Soziologie. *Soziale Welt*, 29(2), 129-154.
- Opp, K. D. (2019). Die Theorie rationalen Handelns, das Modell der Frame-Selektion und die Wirkungen von Bestrafungen auf Kooperation. *Zeitschrift für Soziologie*, 48(2), 97-115. doi: 10.1515/zfsoz-2019-0008
- Ostrom, E. (2007). Institutional rational choice. In P. A. Sabatier (Hrsg.), *Theories of the policy process* (2. Auflage, S. 21-64). Cambridge, MA: Westview Press.
- Pappi, F. U. & Shikano, S. (2004). *Untersuchung Gesundheitssystem* (Studiennummer 4330). Köln: Zentralarchiv für empirische Sozialforschung.
- Parsons, T. (1937). *The Structure of Social Action*. New York, NY: McGrawHill.
- Parsons, T., (1951). *The Social System*. Routledge & Kegan Paul.
- Parsons, T. (1969). Evolutionäre Universalien der Gesellschaft. In W. Zapf (Hrsg.), *Theorien des sozialen Wandels* (S. 55-74). Köln/Berlin: Kiepenheuer & Witsch.
- Peters, B. G. (2019, 4. Aufl.). *Institutional theory in political science: The new institutionalism*. Cheltenham, GB/Northampton, MA: Edward Elgar Publishing.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In R. E. Petty & J. T. Cacioppo (Hrsg.), *Communication and persuasion* (S. 1-24). New York, NY: Springer. doi: 10.1007/978-1-4612-4964-1
- Powell, W. W., & DiMaggio, P. J. (Hrsg.). (1991). *The new institutionalism in organizational analysis*. London, GB/Chicago, IL: University of Chicago press.
- Ragin, C., & Zaret, D. (1983). Theory and method in comparative research: Two strategies. *Social forces*, 61(3), 731-754. doi: 10.1093/sf/61.3.731
- Rotter, J. B. (1954). General Principles for a Social Learning Framework of Personality Study. In J. B. Rotter (Hrsg.), *Social learning and clinical psychology* (S. 82-104). Englewood Cliffs, NJ: Prentice-Hall. doi:10.1037/10788-004
- Sauermann, J. (2020). On the instability of majority decision-making: testing the implications of the ‘chaos theorems’ in a laboratory experiment. *Theory and Decision*, 1-22.

- Sauermann, J. (2018). Do individuals value distributional fairness? How inequality affects majority decisions. *Political Behavior*, 40(4), 809-829. doi: 10.1007/s11109-017-9424-6
- Sauermann, J., & Kaiser, A. (2010). Taking Others into Account: Self-Interest and Fairness in Majority Decision Making. *American Journal of Political Science*, 54(3), 667-685. doi: 10.1111/j.1540-5907.2010.00453.x
- Schmidt, V. A. (2008). Discursive institutionalism: The explanatory power of ideas and discourse. *Annu. Rev. Polit. Sci.*, 11, 303-326. doi: 10.1146/annurev.polisci.11.060606.135342
- Schmidt, V. A. (2011). Reconciling ideas and institutions through discursive institutionalism. In D. Béland & R. H. Cox (Hrsg.), *Ideas and politics in social science research*, 47-64. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199736430.003.0003
- Schmitt, M. (1990). Die Konsistenzkontroverse. In D. Albert, K. Pawlik, K.-H. Stapf, & W. Stroebe (Hrsg.), *Lehr- und Forschungstexte Psychologie 36. Konsistenz als Persönlichkeitseigenschaft? Moderatorvariablen in der Persönlichkeits- und Einstellungsforschung* (S. 14-28). Berlin/Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-75998-7_3
- Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In H. Sauermann (Hrsg.), *Beiträge zur experimentellen Wirtschaftsforschung* (S. 136-168). Tübingen: J.C.B. Mohr (Paul Siebeck).
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99-118. doi: 10.2307/1884852
- Six, B., & Eckes, T. (1996). Metaanalysen in der Einstellungs-Verhaltens-Forschung. *Zeitschrift für Sozialpsychologie*, 27(1), 7-17.
- Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2), 274-279.
- Stocké, V. (2002). *Framing und Rationalität: Die Bedeutung der Informationsdarstellung für das Entscheidungsverhalten*. München: Oldenbourg.
- Thelen, K. (1999). Historical institutionalism in comparative politics. *Annual review of political science*, 2(1), 369-404. doi: 10.1146/annurev.polisci.2.1.369
- Thelen, K., & Steinmo, S. (1992). In S. Steinmo, K. Thelen, & F. Longstreth (Hrsg.), *Structuring politics. Historical institutionalism in comparative analysis* (S. 1-32). Cambridge (UK): Cambridge University Press.
- Thiery, P. (2019). Moderne politikwissenschaftliche Theorie. In H.-J. Lauth & C. Wagner (Hrsg.), *Politikwissenschaft. Eine Einführung* (9. Auflage, S. 196-234). Paderborn: Schöningh.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297-323.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- Vroom, V. H. (1964). *Work and Motivation*. New York, NY: Wiley.
- Weber, M. (1921). *Wirtschaft und Gesellschaft: Grundriss der verstehenden Soziologie*. Tübingen: Mohr Siebeck.
- Wiswede, G. (1977). *Rollentheorie*. Stuttgart/Berlin/Köln/Mainz: Kohlhammer.

3 Konditionale Kooperation in experimentellen Öffentliches-Gut-Spielen mit und ohne Feedback: Erklärung oder Artefakt?¹

Conditional Cooperation in Experimental Public Goods Games: Explanation or Artefact?

Dirk Betz^{*#§}

^{*} Cologne Center for Comparative Politics (CCCP), University of Cologne, IBW-Gebäude, Herbert-Lewin-Straße 2, 50931 Köln, Germany.

[#] Otto-von-Guericke-Universität (OVGU), Fakultät für Wirtschaftswissenschaft (FWW), Lehrstuhl für Wirtschaftspolitik, Gebäude 22, Universitätsplatz 2, 39106, Magdeburg, Germany. E-Mail: dirk.betz@ovgu.de

[§] GESIS – Leibniz-Institut für Sozialwissenschaften, Unter Sachsenhausen 6-8, 50667 Köln, Germany. E-Mail: dirk.betz@gesis.org (Kontaktadresse)

Für die finanzielle Unterstützung wird GESIS-DAS, SocioCologne e.V. und der Deutschen Forschungsgemeinschaft (DFG) herzlich gedankt. Ein besonderer Dank gebührt Claudia Biniossek, Jeannette Brosig-Koch, Hartmut Esser, B.E. und dem elfe-Labor der Universität Duisburg-Essen für die Unterstützung bei der Durchführung der Experimente sowie Jürgen Friedrichs für seine langjährige Begleitung.

Schlagworte: Rational Choice, Öffentliches-Gut-Spiel, Modell der Frame-Selektion, BRAC-Rahmenmodell, konditionale Kooperation

Keywords: Rational Choice, Public Goods Game, Model of Frame Selection, BRAC framework, Conditional Cooperation

JEL Klassifikation: C71; C9; C91; H41

1 Betz, D. (2020). *Konditionale Kooperation in experimentellen Öffentliches-Gut-Spielen mit und ohne Feedback: Erklärung oder Artefakt?* Manuskript eingereicht zur Publikation bei der Fachzeitschrift *FQS Forum: Qualitative Sozialforschung*

Zusammenfassung

Aktuelle Studien zum experimentellen Öffentliches-Gut-Spiel (ÖGS) identifizieren neben *Trittbrettfahrern* und *konditionalen Kooperatoren* eine Vielzahl weiterer Typen, ohne einen der Untersuchungsstränge des Mixed-Methods-Designs (MMD) für die Typenbildung einzubeziehen. Aus psychologischen und soziologischen Rahmenmodellen ist ableitbar, dass nicht die Spielsituation den Typ definiert, sondern empirisch zu erheben ist, welchen situativ passenden Typ Individuen (aktiv) wählen. Hieraus wird die Forschungsfrage abgeleitet, welchen Einfluss die objektiv vorgegebene Spielstruktur des ÖGS, die spezifische Gruppenkultur und das Feedback auf das Denken und Handeln der Individuen haben. Zur Überprüfung wurde das ÖGS mit Feedback (NoP-Treatment, $n = 16$) und ohne Feedback (NoF-Treatment, $n = 16$) über 30 Runden gespielt. Nach Abschluss der Experimente wurden anhand qualitativer Fragen und quantitativer Skalen die Entscheidungsgründe und die Situationswahrnehmung erhoben. Als Ergebnis können in drei von vier Gruppen gruppenspezifische Gleichgewichte und eine Konvergenz von Strategien und Typen nachgewiesen werden (NoP-Treatment). Im NoF-Treatment zeigen nur fünf der 16 Versuchspersonen ein mit den Annahmen der Rational-Choice-Theorie (RCT) vereinbares Verhalten, von denen zwei als *Trittbrettfahrer* klassifizierbar sind. Zudem sind bei der Mehrheit der Versuchspersonen inkonsistente mentale Prozesse und Verhaltensmuster beobachtbar. Rahmenmodelle entfalten in uneindeutigen Situationen eine große Prognosekraft und können die RCT als einen Spezialfall in sich integrieren.

Abstract

Recent studies on the experimental Public Goods Game (PGG) identify a variety of types in addition to *free-riders* and *conditional cooperators* without including one of the research strands of Mixed Method Design (MMD). Psychological and sociological behavioral framework models assume that the game situation does not define the type. Therefore, it must be empirically ascertained which type individuals (actively) choose in a given situation. We investigate the effect of the objectively given game structure, the specific group culture, and the feedback on the individual decision by conducting the repeated PGG experiment with feedback (NoP-Treatment, $n = 16$) and without feedback (NoF-Treatment, $n = 16$) over thirty rounds. Thereafter, qualitative questions and quantitative scales were used to determine the decision reasons and the perception of the situation. As a result, group-specific equilibria and convergence of strategies and types can be demonstrated in three of four groups (NoP-Treatment). In the NoF-Treatment, only five of the 16 participants show behavior compatible with assumptions of rational choice theory (RCT), two of them classified as *free riders*. In addition, inconsistent mental processes and behavioral patterns can be observed in the majority of the participants. In our study, behavioral framework models demonstrate a high predictive power in ambiguous situations and can integrate RCT as a special case.

Resumen

En estudios recientes sobre el Juego de Bienes Públicos (Public Goods Game: PGG) en la economía experimental se identifican otros tipos además de los *free-riders* y los *cooperadores condicionales*, sin incluir una de las líneas de investigación del Diseño de Método Mixto (MMD) para una construcción empírica de tipos. Los modelos de framework de comportamiento psicológico y sociológico asumen que no es la situación de juego la que define el tipo. Por el contrario, debe determinarse empíricamente qué tipo eligen los individuos (activamente) en una situación determinada. Examinamos el efecto de la estructura del juego, la cultura específica del grupo y la retroalimentación en la toma de decisiones individuales. Para probarlo, se jugó el PGG con feedback (NoP-Tratamiento, $n = 16$) y sin feedback (NoF-Tratamiento, $n = 16$) durante 30 rondas. Una vez finalizados los experimentos, se utilizaron preguntas cualitativas y escalas cuantitativas para determinar empíricamente las razones de la decisión y la percepción de la situación. Como resultado, los equilibrios específicos de los grupos y la convergencia de las estrategias y los tipos pueden demostrarse en tres de los cuatro grupos (NoP-Tratamiento). En el tratamiento NoF, sólo cinco de los 16 participantes muestran un comportamiento compatible con los supuestos de la teoría de la selección racional (RCT), dos de ellos pueden ser clasificados como *free-riders*. Además, en la mayoría de los sujetos de prueba se observan procesos mentales y pautas de comportamiento inconsistentes. Los modelos de framework y comportamiento despliegan un gran poder de predicción en situaciones ambiguas y pueden integrar el RCT como un caso especial.

Einleitung

Konditionale Kooperation als Erklärung der Ergebnisse im experimentellen Öffentliches-Gut-Spiel

2007 publizierten die Verhaltensökonomern Ernst Fehr und Herbert Gintis einen Grundlagenartikel in der *Annual Review of Sociology*, der die Erkenntnisse der Verhaltensökonomie zu experimentellen Kollektivgutspielen vor dem Hintergrund ihres *Beliefs, Preferences, and Constraints*-Ansatzes einer breiten sozialwissenschaftlichen Öffentlichkeit zugänglich machte. In der Folgezeit erlangte insbesondere der bereits im Jahr 2000 veröffentlichte Artikel von Fehr und Gächter (im Folgenden: F&G) eine große Bedeutung in der Politikwissenschaft, Soziologie und anderen Sozialwissenschaften, da dieser anhand von Varianten eines experimentellen Öffentliches-Gut-Spiels (ÖGS) mit und ohne Bestrafungsoption Abweichungen vom Eigennutzmaximierungsprinzip und von engen Varianten der Rational-Choice-Theorie (RCT) nachweisen konnte. Grundidee des experimentellen wiederholten ÖGS ist, dass die Spieler² einer Gruppe in jeder Runde mit einer Geldsumme ausgestattet werden, die sie entweder in ein gemeinsames Projekt einzahlen oder behalten können. Die Gemeinschaftskasse wird anschließend um einen vorher festgelegten Faktor, den MPCR (marginal per capita return), vermehrt und gleichmäßig an alle Spieler ausbezahlt, unabhängig davon, ob diese zum gemeinsamen Projekt beigetragen haben oder nicht. Kollektiv stellt sich die Gruppe am besten, wenn alle Spieler ihr gesamtes Geld in das gemeinsame Projekt einzahlen (Pareto-Optimum). Individuell erhält ein Spieler in einer Runde die höchste Auszahlung, wenn alle anderen Spieler ihr gesamtes zur Verfügung stehendes Geld in das Gemeinschaftsprojekt einzahlen, er selbst jedoch keinen Beitrag leistet und seine gesamte Geldausstattung behält (Trittbrettfahrerverhalten). Die Ökonomische Nutzentheorie und enge Varianten der RCT gehen von rationalem, eigennutzmaximierendem Verhalten aus und prognostizieren daher für das (endliche) anonyme ÖGS, dass als dominante Strategie kein Spieler zum öffentlichen Gut beiträgt (Nash-Gleichgewicht).

Experimentell wurden für das wiederholte ÖGS ohne Bestrafung von F&G (2000) mittlere Beiträge in der ersten Runde und ein Sinken der Beiträge in den Folgerunden nachgewiesen. Die Befunde konnten sich in Folgestudien als robust erweisen (Thöni & Volk, 2018; Zelmer, 2003). Durch Einführung der Option, unkooperative Mitspielende zu bestrafen, konnten F&G (2000) für das wiederholte ÖGS ein höheres Anfangsniveau für Beiträge in der ersten Runde und weiter steigende Beiträge in den Folgerunden nachweisen, die sich auf sehr hohem Niveau stabilisieren. Besteht eine Bestrafungsoption, werden *Trittbrettfahrer*, die in einer Runde wenig beitragen, anders als von der RCT prognostiziert, oftmals von den Mitspielenden bestraft, auch wenn die Bestrafung Geld kostet. Die von F&G (2000) nachgewiesenen Abweichungen von den Prognosen der ökonomischen Nutzentheorie haben in der Folgezeit in

2 Bezeichnungen, die in der männlichen Form verwendet werden, schließen eine adäquate Sprachform für weiblich und divers ein.

den Wirtschafts- und Sozialwissenschaften zu einer Erweiterung des Nutzenbegriffs um soziale Präferenzen sowie zur wachsenden Verbreitung erweiterter RCT-Varianten beigetragen.

Zur Erklärung der experimentellen Befunde identifizieren Fischbacher, Gächter und Fehr (im Folgenden: FGF) (2001) anhand der Strategiemethode (Selten, 1967; Mitzkewitz & Nagel, 1993) unterschiedliche Verhaltenstypen. Hierbei ist konditionale Kooperation definiert als die Orientierung der eigenen Beiträge zum Gemeinschaftsgut an den Beiträgen der anderen Mitspielenden: „people who are willing to contribute more to a public good the more others contribute“ (FGF, 2001, S. 397). Mit etwa 50 % sind *konditionale Kooperateure* der am weitesten verbreitete Typ, dessen eigener Beitrag zum Kollektivgut proportional zum Durchschnitt der Beiträge der anderen Gruppenmitglieder ansteigt bzw. fällt. Nach FGF (2001) kann ein konditional kooperierender Spielertyp als intrinsisch motiviert oder als durch soziale Präferenzen wie Altruismus, *warm glow of giving* (Andreoni, 1990), Ungleichheitsaversion oder Reziprozität geleitet interpretiert werden. Etwa ein Drittel der Spieler (30 %) kann als *Trittbrettfahrer* (*Free Rider*) klassifiziert werden und leistet keine Beiträge zum Kollektivgut. Weitere 14 % der Spieler leisten zunächst, wie die *konditionalen Kooperateure*, mit steigendem Gruppendurchschnitt auch höhere eigene Beiträge zum Kollektivgut. Steigen die durchschnittlichen Beiträge der anderen Gruppenmitglieder jedoch über 50 %, brechen die Beiträge dieses Spieler-Typs ein und nähern sich Null, je mehr die anderen Gruppenmitglieder durchschnittlich zum Kollektivgut beitragen. Dieser dritte identifizierte Typ wird als *hump-shaped* (FGF, 2001) bzw. *triangle contributors* (Fischbacher & Gächter, 2010) bezeichnet. Insgesamt können damit 94 % der Spieler einem dieser drei Typen zugeordnet werden. Das auch auf Aggregatebene beobachtbare Beitragsverhalten der Spieler ist auf Individualebene durch den hohen Anteil des konditional kooperierenden Spielertyps erklärbar. Die Argumentation von FGF (2001) ist hierbei, dass das Absinken der Kooperation hauptsächlich durch das Zusammenspiel zweier Faktoren bedingt ist: *Trittbrettfahrer* in einer Gruppe, die nichts zum Gemeinschaftsgut beitragen, und *konditionale Kooperateure*, die mit dem unkooperativen Verhalten der *Trittbrettfahrer* konfrontiert sind und hierauf reziprok mit eigenen sinkenden Beiträgen antworten, was aus Sicht der Autoren zusätzlich durch einen Bias des konditional kooperierenden Typs in Richtung Eigennutzmaximierung unterstützt wird. Diese Erklärungsansätze mittlerer Reichweite für das ÖGS sind in ihrem Geltungsbereich jedoch beschränkt auf Situationen, in denen aus subjektiver Perspektive des Spielers Kooperation im Vordergrund steht (prosoziales Verhalten bei Kooperationsspielen sowie in Spieldesigns mit kooperativem Framing), während in Marktspielen sowie wettbewerbsorientiertem Framing prosoziales Verhalten kaum beobachtbar ist (Lieberman, Samuels, & Ross, 2004). Aus dem Bedarf nach einem übergeordneten, gemeinsamen Erklärungsansatz zu der Frage, wie die widersprüchliche Empirie für verschiedene Spielklassen (Kooperationsspiele, Marktspiele etc.) in einen einheitlichen theoretischen Rahmen integriert werden kann, erlangten in der Verhaltensökonomie zum Zeitpunkt der Publikation von FGF (2001) Modelle der Ungleichheitsaversion immer größere Bedeutung (z. B. Fehr & Schmidt, 1999; ERC-Modell von Bolton & Ockenfels,

2000), die das Verhalten sowohl von *konditionalen Kooperatoren* als auch von *Trittbrettfahrern* erklären können. Sauermann und Kaiser (2010) finden in ihrer experimentellen Studie zu (politischen) Mehrheitsentscheidungen ebenfalls empirische Evidenz, die die Ergebnisse von FGF (2001) bezüglich ihrer Erklärung mit Modellen der Ungleichheitsaversion stützt. Sauermann (2018) konnte in seiner experimentellen Studie zu (politischen) Mehrheitsentscheidungen dagegen Auszahlungsmaximierung und neben Ungleichheitsaversion auch noch Verteilungsgerechtigkeit (distributional fairness) als weitere soziale Präferenz empirisch belegen.

Fischbacher und Gächter erweitern 2010 das experimentelle Design von FGF (2001) und postulieren nun, dass die Beiträge zum Kollektivgut letztlich deshalb sinken, weil die Spieler *unperfekte konditionale Kooperatoren* sind, die sich im Zeitverlauf über mehrere Runden schließlich wie *Trittbrettfahrer* verhalten und somit in Richtung des auszahlungsmaximierenden Typs im Sinne von *Selfish in the End* (Brosig-Koch, Riechmann, & Weimann, 2007) konvergieren, auch wenn nur eine Minderheit unter ihnen durch bloßen Eigennutz motiviert ist. Die experimentellen Befunde von FGF (2001) zum wiederholten ÖGS konnten sich in der Folgezeit vor dem Hintergrund wachsender empirischer Evidenz als robust und reliabel erweisen. Neben der FGF-Typologie werden in der Literatur auch Typologie-Varianten und Erweiterungen genutzt, die teils erheblich voneinander abweichen. In diesem Kontext diskutieren etwa Thöni und Volk (2018) den empirisch beobachtbaren Typ des *unkonditionalen Kooperators*. Einige dieser Typologien wurden hierbei entwickelt, ohne auf ein einheitliches Verfahren zur Definition und Evaluation von Typen zurückzugreifen. Um diesen Mangel zu überwinden, evaluieren Fallucchi, Luccasen und Turocy (2019) anhand von Machine Learning-Methoden die Erklärungskraft unterschiedlicher Typologien für das Verhalten im ÖGS. Anhand von Gütekriterien wird evaluiert, zu welchem Grad Spieler des gleichen Typs ein ähnliches Entscheidungsverhalten zeigen und zu welchem Grad zwischen unterschiedlich typisierten Spielern Unterschiede im Verhalten bestehen. In ihrer Metaanalyse finden die Autoren eine Typologie der Spieler, die von FGF (2001) abweicht: Eigennutzmaximierer (*own maximisers*), bei denen der modale Beitrag null beträgt, *strong conditional cooperators*, die jeweils exakt den Durchschnittsbeitrag leisten; *schwach-konditionale Kooperatoren*, die generell steigende Beitragsstrategien verfolgen, aber nicht den exakten Durchschnittsbeitrag leisten, sowie *Sonstige*, unter die auch Spieler fallen, die in jeder Runde maximal beitragen. Diese Vierer-Typologie wird von den Autoren im Verlauf ihres Artikels durch Differenzierung der Gruppe *Sonstige* zu einer Fünfer-Typologie erweitert, indem der Typ *unconditional high contributor* (UCH) hinzugefügt wird, der in ihrer Studie 4.7 % der Versuchspersonen (Vpn) ausmacht. Während diese UCH darüber definiert sind, immer alles beizutragen, definieren Thöni und Volk (2018) Spieler als *unkonditionale Kooperatoren*, die unabhängig vom Beitragsverhalten der Anderen immer einen konstanten Beitrag zum Kollektivgut leisten. Da Fallucchi et al. (2019) sowie Thöni und Volk (2018) einen theoriefernen, empiristischen Ansatz verfolgen, werden weder die hinter einem Verhaltenstyp stehenden mentalen Mechanismen, noch gruppenspezifische Ausprägungsformen oder die intertemporale und intersituative Konsistenz der Typen betrachtet.

So weit – so gut?

Trotz der Robustheit der Befunde zum mehrheitlich kooperativen Verhalten von Spielern bleiben einige Befunde bis dato ohne zufriedenstellende Erklärung. Im Folgenden sollen daher zuerst einige methodische Überlegungen diskutiert werden, die sich aus der Tatsache ergeben, dass die bisherigen Arbeiten hauptsächlich die Aggregatenebene betrachten und adressieren, jedoch nur wenige Betrachtungen auf der Meso- und Mikroebene anstellen. Daran anschließend werden empirische Befunde vorgestellt, die auch weiten RCT-Ansätzen zu widersprechen scheinen, und die hierzu in der Literatur verfolgten theoretischen Lösungsansätze dargestellt. Danach werden alternative, handlungstheoretisch orientierte Erklärungsmechanismen aus Psychologie, Soziologie und Neurowissenschaften skizziert, die sich teilweise nicht mehr der RCT und ihren grundlegenden Annahmen und Mechanismen zuordnen lassen und sich möglicherweise als leistungsfähige theoretische Rahmenkonzepte für die Erklärung des ÖGS anbieten.

Einige methodische Überlegungen

Strukturell verzögertes Feedback

Im Stranger Design ist ein Feedback objektiv irrelevant, weil in jeder Runde mit anderen Mitspielenden gespielt wird. Demgegenüber ist das Feedback im Partner Design objektiv relevant. Jedoch besteht hier für die gesendeten Signale ein strukturell bedingter Timelag. Wenn Spieler A seinen Mitspielenden die Änderung seiner Strategie signalisieren möchte, also beispielsweise ein Spieler, der in der Vorrunde t_0 zunächst keine Beiträge leistet, in der Runde t_1 ein sehr positives Signal³ in Form hoher eigener Beiträge sendet, empfangen seine Mitspielenden dieses Signal erst in der Folgerunde t_2 und eine reziproke Verhaltensantwort der Mitspielenden, die diese in t_2 zurücksenden, ist damit frühestens in der darauf folgenden Runde t_3 als konditional kooperative Reaktion für alle Spieler erkennbar. Spieler A bekommt also frühestens in t_3 die Rückmeldung, ob auch seine Mitspielenden konditional kooperieren und kann dieses in t_3 erworbene Wissen erst ab t_4 in seine Beiträge zum gemeinsamen Projekt einspeisen. Unmittelbar konditional kooperatives Verhalten ist damit im wiederholten ÖGS strukturell nicht möglich. Im ÖGS spielt diese zeitliche Verzögerung (Timelag) eine wichtige Rolle, da hohe Beiträge eines Spielers von den übrigen Gruppenmitgliedern nicht direkt positiv wertgeschätzt werden (können). Rationale Akteure wissen, dass sich in einer Gruppe sowohl *Konditionale Kooperative* mit rationalen Erwartungen als auch *Konditionale Kooperative* mit adaptiven Erwartungen befinden können. Rationale Akteure müssen daher die strukturelle Timelag-Falle des ÖGS erkennen, die bei *konditionalen Kooperatoren* mit adaptiven,

3 Clark und Sefton (2001) zeigen, dass im wiederholten ÖGS das Partner Design die Möglichkeit zum *Signaling* bietet und Kooperation nicht nur das Ergebnis von Lernen ist. Ackermann und Murphy (2019) finden demgegenüber jedoch keine Anhaltspunkte für systematisches Signaling.

nicht rationalen Erwartungen zu antizyklischen Mustern und einem strukturellen *Schweinezyklus-Effekt* (Hanau, 1928) bzw. *Cobweb-Effekt* (Ezekiel, 1938) führt. Falls sich in diesem Szenario auch *Trittbrettfahrer* in der Gruppe befinden, ergeben sich im Aggregat zudem kontinuierlich sinkende Beiträge. Intendieren Spieler hohe Kooperationsraten ihrer Gesamtgruppe, ist daher konditionale Kooperation keine geeignete Verhaltensstrategie, wohl aber das Durchhalten hoher Kooperationsraten für mindestens zwei Folgerunden. Bereits Axelrod (1984) hat in seinen Simulationen zur dynamischen Evolution der Kooperation auf die Möglichkeit hingewiesen, dass in bestimmten Umwelten nicht eine harte Tit-for-Tat-Strategie, sondern eine nachsichtigere Tit-for-two-Tat-Strategie erfolgreicher sein kann. Vor diesem Hintergrund stellt sich die bisher unbeantwortete Frage, zu welchem Grad das durchschnittliche Abfallen der Kooperationsraten im wiederholten ÖGS unter Umständen lediglich ein methodeninduziertes Artefakt darstellt, das durch das verzögerte Feedback bedingt ist.

Split des Treatments durch gruppenspezifische Beiträge ab der zweiten Runde

In wiederholten ÖGS erhalten Versuchspersonen erst ab der zweiten Runde Informationen über das Entscheidungsverhalten ihrer Mitspielenden und damit die Möglichkeit, ihre Erwartungen entsprechend anzupassen. Die Vpn befinden sich somit in der ersten und der zweiten Runde in völlig unterschiedlichen (Spiel-)Situationen (Fischbacher & Gächter, 2010, S. 548). Erkennbare (theoretische) Konsequenzen werden von den Autoren aus dieser Tatsache jedoch nicht gezogen, beispielsweise, dass Vpn in der ersten Runde möglicherweise zunächst ein defensives Orientierungsverhalten zeigen, um das Verhalten der übrigen Gruppenmitglieder abzuwarten. Esser (2018) argumentiert hierzu, dass die Spieler in der ersten Runde weitgehend nach ihren privaten Motiven handeln, und nimmt eine „Gleichverteilung der beiden Frames (EF und AF) über die Akteure“ (S. 17) an. In der ersten Runde des ÖGS finden sich jeweils etwa zur Hälfte „Altruisten mit einem Altruismus-Frame (AF-Typ)“ (S. 17), die mindestens die Hälfte ihrer Ausstattung beitragen. Die andere Hälfte der Spieler sind „Egoisten mit einem Egoismus-Frame (EF-Typ)“ (S. 17), die weniger als die Hälfte in das gemeinsame Projekt einzahlen. Das Verhalten der Spieler „definiert“ dann die Situation für die 2. Runde (und danach sukzessive) immer deutlicher“ (S.17) und führt in den Folgerunden insbesondere beim AF-Typ zu Anpassungsprozessen und zu einem zunehmenden Verfall der Kooperation in der Gesamtgruppe.

In experimentellen ÖGS mit mehreren Runden erhalten Vpn in sehr kooperativen Gruppen ab der zweiten Runde ein anderes Feedback als Vpn in unkooperativen Gruppen. Vpn mit positivem Feedback sind einem anderen Stimulus und damit einer anderen unabhängigen Variablen ausgesetzt als Vpn, die in ihrer unkooperativen Gruppe ein negatives Feedback erhalten. Sobald nicht alle Gruppen in der vorangegangenen Runde denselben kollektiven Beitrag geleistet haben und somit dasselbe Feedback erhalten, ergeben sich trotz identischem Spieldesign aufgrund der unterschiedlichen Reaktion der Mitspielenden aus dem gleichen ÖGS-Treatment ab der zweiten Runde mehrere unterschiedliche unabhängige Variablen, die folglich

auch in der Datenanalyse zu trennen sind. Dies entspricht auch den Ergebnissen von Ackermann und Murphy (2019) die zeigen, dass der durchschnittliche Beitrag der Gesamtgruppe in der ersten Runde sowie die Tendenz der Gruppenmitglieder, konditional zu kooperieren, einen starken Einfluss auf das Gesamtbeitragsniveau haben. In der Literatur wird häufig übersehen, dass die subjektiv wahrgenommene Spielstruktur im wiederholten ÖGS somit ab der zweiten Runde nicht nur aus dem vom Experimentator vorgegebenen Spieldesign, sondern auch aus den Reaktionen der Mitspielenden resultiert.

Umgang mit widersprüchlicher beziehungsweise ungeklärter Empirie

Restart-Effekt

Im wiederholten ÖGS kommt es nicht nur nach der Einführung einer Bestrafungsoption zu einem Anstieg der Beiträge, sondern auch bei erneutem Start des Experimentes ohne Bestrafungsoption ist ein Anstieg der Beiträge (isolierter Restart-Effekt) zu beobachten (Chaudhuri, 2018; Esser, 2018). Ist eine Vpn konditional an den Beiträgen ihrer Mitspielenden orientiert, müssten theoretisch ihre Beiträge nach einem Restart weiter sinken statt zu steigen, falls sich die anderen in der unmittelbar vorangegangenen Runde unkooperativ verhalten haben. Empirisch ist dies vor allem im Partner Design (Chaudhuri, 2018) nicht der Fall. Somit führen objektiv relevante Designunterschiede nur beim Restart, nicht jedoch zwischen den Runden zu einem empirisch unterschiedlichen Entscheidungsverhalten. Eine differenzierte Analyse der Messwerte auf kleinerer Granularität, das heißt auf Gruppen- und auf Individual-ebene, erscheint daher notwendig um zu klären, ob der empirisch beobachtbare isolierte Restart-Effekt beispielsweise dadurch zu erklären ist, dass konditional kooperierende Akteure (i) spezifische, bisher nicht berücksichtigte Erwartungen bezüglich des Neustarts haben oder (ii) weiteren Motiven folgen. Änderungen der Erwartungen könnten beispielsweise aus einem erhöhten Optimismus der Gruppenmitglieder resultieren (Chaudhuri, 2018) oder sich aus dem Motiv ergeben, sich wohlwollend und nicht-konditional kooperierend zu verhalten, um der Gruppe nochmals eine Chance auf höhere, gruppeneffizientere Gleichgewichte zu geben.

Empirismus

FGF (2001) konnten drei Verhaltenstypen im ÖGS empirisch nachweisen: *Konditionale Kooperateure*, Vpn mit *hump-shaped* Beiträgen und *Trittbrettfahrer*. Die verbleibenden, teils sehr heterogenen Beitragsmuster wurden von den Autoren unter dem Begriff *Andere* subsumiert. *Trittbrettfahrer* entsprechen hierbei den Prognosen einer engen RCT, sowohl *konditionale Kooperateure* als auch der *hump-shaped* Typ lassen sich durch Erweiterungen der RCT wie beispielsweise die *Theory of Equity, Reciprocity, and Competition* (ERC-Modell) von Bolton und Ockenfels (2000) oder die Theorie der Ungleichheitsaversion (Fehr & Schmidt, 1999) erklären, die ab der Jahrtausendwende weite Verbreitung fanden. In der Folgezeit wurde in der Literatur zunehmend

von Befunden berichtet, die von der FGF-Typologie nicht mehr ausreichend erklärt werden können. Entgegen Theorien der Ungleichheitsaversion ergänzen Thöni und Volk (2018) die FGF-Typologie noch um den zusätzlichen Typ *unkonditionale Kooperationsrateure*, jedoch ohne ein transparentes Verfahren zur Definition und Evaluation von Typen und deren theoretische Verortung zu bieten. Fallucchi et al. (2019) brechen mit dieser Vorgehensweise, indem sie sich von einer theoriegeleiteten Erforschung der Wirkmechanismen explizit abwenden. Die Autoren ergänzen wie oben bereits dargestellt das ÖGS um den Typen *unconditional high contributor* (UCH) indem sie Typen anhand deren Trennschärfe zur Identifizierung von Spielern mit ähnlichen Entscheidungsmustern evaluieren. Der Entdeckung immer neuer Typen steht das Phänomen gegenüber, dass der Anteil der Typen zwischen Experimenten stark variiert; beispielsweise haben FGF (2001) in ihrer Untersuchung keine *unkonditionalen Kooperationsrateure* gefunden. Damit steht eine der Kernannahmen des ökonomischen Verhaltensparadigmas zur Disposition, dass Typen situationsübergreifende, temporal stabile Persönlichkeitszüge und damit eine exogene Variable darstellen. Die Identifizierung neuer Typen im ÖGS wird begleitet durch die Ergänzung bestehender Ansätze um theoretische Konzepte geringer Reichweite, beispielsweise eine aus Immanuel Kants kategorischem Imperativ abgeleitete Meta-Gerechtigkeitsnorm (Alger & Weibull, 2013, 2019), aber auch Ansätze mit Bezug zu Lerntheorien (z. B. Erev & Haruvy, 2015). Einer weiteren Möglichkeit, die heterogenen empirischen Befunde mit einem theoretischen Ansatz zu erklären, wird in Theorienvergleichen beispielsweise bei Mietten, Kosfeld, Fehr, & Weibull (2020) nachgegangen. Zusammenfassend lässt sich nach Hartmut Esser (Biniossek, 2019, S. 9) eine zunehmende Hinwendung zu einer theorielosen Variablen-Forschung beobachten. Vor diesem Hintergrund wird in der vorliegenden Untersuchung versucht, die empirischen Befunde in ein übergreifendes theoretisches Paradigma zu integrieren.

Limitiertes Feedback

Im wiederholten ÖGS liegen in der ersten Runde noch keine Informationen über die Entscheidungen der Mitspielenden vor, da erst ab der zweiten Runde ein Feedback gegeben wird. Fischbacher und Gächter (2010) argumentieren, dass die Versuchspersonen mit ganz unterschiedlichen Prädispositionen in die Experimentalsituation hineingehen und bezeichnen diese als „intuitive (*home-grown*) beliefs about others' contributions“ (S. 548), die die Grundlage von Entscheidungen ohne Feedback in der ersten Runde bilden. Aus dieser Argumentation lässt sich ableiten, dass diese *home-grown beliefs* in Treatments mit limitiertem Feedback oder ohne Feedback auch in den folgenden Runden eine zentrale Rolle spielen. Informationen über die individuellen und kollektiven Beiträge (z. B. Spendenbeiträge, geleistete Steuern zur Erstellung öffentlicher Güter etc.) variieren im Alltag stark, je nachdem um welches Kollektivgut es sich handelt. Bei einer Übertragung experimenteller Befunde zur Wirkung von Feedback im ÖGS auf reale öffentliche Güter ist daher zu beachten, dass es im Alltag oftmals kaum möglich ist zu erfahren, wer welche monetären und nicht-monetären Beiträge geleistet hat (z. B. Verzicht auf Flugreisen zum Wohle der Umwelt), welche

Auswirkungen der geleistete Beitrag auf die Erstellung des öffentlichen Gutes hat (z. B. Spenden für eine Entwicklungshilfeorganisation zur Bekämpfung von Hunger), wie hoch das geleistete (relative) Opfer im Verhältnis zu Einkommen und Vermögen ist und wie stark der Beitragleistende (Steuerzahler, Spender, Clubmitglied etc.) und sein persönliches soziales Umfeld von der Erstellung des öffentlichen Gutes profitieren. Die Literatur zu limitiertem Feedback fokussiert hauptsächlich auf die Einschränkung der Information, wer in welchem Umfang zum gemeinsamen Projekt beigetragen hat, sowie auf die Darstellungsform von Beiträgen entweder als Einnahmen (earnings) oder als Beitrag (contribution). Nikiforakis (2010) weist hierzu für das wiederholte ÖGS mit Bestrafungsoption bezüglich der Art und Weise der Informationsbereitstellung nach, dass Informationen über das Einkommen (earnings) der anderen Gruppenmitglieder zu niedrigeren Beiträgen führen als im Standarddesign, in dem Informationen über die Beiträge (contributions) gegeben werden. Demgegenüber findet Weimann (1994) heraus, dass die Information über das Verhalten der anderen Spieler keine Auswirkungen auf das individuelle Verhalten hat, was bedeuten würde, dass sich auch ohne Feedback das Verhalten der einzelnen Vpn nicht ändern sollte. Nur wenn die übrigen (fiktiven) Gruppenmitglieder erheblich weniger als die Vpn geben, ändert diese ihr Verhalten und senkt ebenfalls ihre Beiträge. Croson (2001) zeigt, dass die Informationen über individuelle Beiträge der Teammitglieder versus Informationen über Beiträge auf Teamebene zu keinem signifikanten Unterschied im Beitragsverhalten führen. Bigoni und Suetens (2012) zeigen, dass die Aufnahme der Information des Feedbacks im Wesentlichen davon abhängt, welche Art Beitragstyp die jeweilige Vpn ist. So suchen *Trittbrettfahrer* eher nach Informationen, die ihr eigenes Verhalten bestätigen, beispielsweise Informationen über Einnahmen statt über Beiträge. Samak und Sheremeta (2013) weisen auf die Lenkungsfunction des Feedbacks bezüglich sozialer Anerkennung hin. Sie zeigen, dass Beiträge zum öffentlichen Gut signifikant steigen, wenn die Beitragenden soziale Anerkennung erfahren. Vesterlund (2006) geht auf die Rolle von Feedback im Anwendungsbereich von Spenden ein, die als öffentliches Gut interpretierbar sind. Analog zu *konditionalen Kooperatoren* im experimentellen ÖGS erhöht sowohl das Feedback über die Spenden der anderen als auch die Information über die Einzelspende einer Person (*Leader*) die eigenen Spendenbeiträge. Vesterlund (2006) führt hier als Verhaltensmotiv Reziprozität an. Varian (1994) weist demgegenüber darauf hin, dass in einem sequentiellen ÖGS, also mit Feedback über das Verhalten des ersten Spielers, weniger zum öffentlichen Gut beigetragen wird, als wenn alle Vpn gleichzeitig beitragen. Ein Feedback führt in seiner Argumentation somit nicht zu einem Steigen, sondern zum Sinken der Beiträge. Neugebauer, Perote, Schmidt und Loss (2009) untersuchen in einem ÖGS mit drei Personen das empirisch festgestellte Sinken der Beiträge anhand ihres experimentellen Designs mit (INFO) und ohne Feedback (NoINFO) über die Beiträge der anderen Gruppenmitglieder. Sie finden in beiden Treatments (INFO und NoINFO) eine hohe Korrelation zwischen dem eigenen Beitrag zum gemeinsamen Projekt im ÖGS und den Erwartungen bezüglich der Beiträge der anderen Gruppenmitglieder. Weiterhin liegen bei Neugebauer et al. (2009) die Beiträge im NoINFO Treatment signifikant über den Beiträgen im INFO Treatment mit Feedback. Neu-

gebauer et al. (2009) finden somit konditionale Kooperation auch dann, wenn kein Feedback gegeben wird, wobei die Annahmen über die Beiträge der anderen Spieler durchschnittlich immer etwas über den eigenen tatsächlichen Beiträgen liegen. Dies erklärt laut Neugebauer et al. (2009) auch das Sinken des Beitragsniveaus in ÖGS.

Weber (2003) zeigt in einer Reihe von Experimenten, dass abweichend von üblichen Lerntheorien, die von Lernen durch Feedback ausgehen, Lernen auch durch Wiederholung und Erfahrung stattfinden kann. Weber (2003) postuliert, dass es auch ohne Feedback zu einem Nash-Gleichgewicht kommen kann. In seinen Experimenten kommt es entsprechend seiner Vorhersagen auch in seinem Treatment ohne Feedback zu einem Absinken der Beiträge, was Weber (2003) auf Lernprozesse zurückführt. Erev und Haruvy (2015) stützen diese Argumentation zusätzlich, indem sie folgendermaßen argumentieren: „The economic incentives determine the agents’ experience, and this experience in turn drives future behavior“ (S. 638). Laut Erev und Haruvy (2015) kann es somit in oft wiederholten Entscheidungen mit unerheblichen ökonomischen Konsequenzen (*small decisions*) auch ohne Kenntnis der Spielstruktur oder Feedback zu ökonomisch sinnvollen Entscheidungen kommen. Nax, Burton-Chellew, West und Young (2016) zeigen empirisch, dass es auch ohne Information über die Spielstruktur und ohne Feedback über das Verhalten der Mitspielenden aufgrund adaptiven Lernens (basierend auf den eigenen Auszahlungen) zum Nash-Gleichgewicht kommt. Aus methodischer Sicht ist an dieser Argumentation zu kritisieren, dass aus den gegebenen Informationen über die eigenen Auszahlungen bereits konkrete Rückschlüsse auf die Spielsituation, die Spielstruktur und das Verhalten der anderen Spieler gezogen werden können, also das Verhalten auf einer rationalen Evaluation der Spielsituation beruhen kann. Bicchieri und Xiao (2009) zeigen, dass in Spielen ohne Feedback für kooperatives Verhalten hauptsächlich ausschlaggebend ist, ob der jeweilige Spieler erwartet, dass die Mitspielenden die Norm *Kooperation* befolgen. Bicchieri und Xiao (2009) erklären dazu, dass die Befolgung sozialer Normen hauptsächlich davon abhängt, ob der Spieler erwartet, dass eine Mehrheit der relevanten Anderen diese Norm befolgt (*empirical expectation*) und nicht so sehr davon, dass andere erwarten, dass man sich an diese Norm hält (*normative expectation*). Zusammenfassend hat empirisch das Ausbleiben von Feedback teils erwartete, teils keine und teils inverse Effekte auf die Höhe der individuellen sowie der aggregierten Beiträge zum öffentlichen Gut.

Leistungsfähigkeit alternativer Entscheidungs- und Verhaltenstheorien

Mit den bisherigen engen und weiten RCT-Erklärungsansätzen ist es auch unter Einbeziehung sozialer Präferenzen, subjektiver Erwartungen und der Klassifizierung oben dargestellter zusätzlicher Spielertypen nicht möglich, die Fülle heterogener Befunde zum (wiederholten) ÖGS umfassend zu erklären. Betrachtet man die Studien auf Gruppen- und auf Individualebene, ergibt sich ein Bild mit zunehmend un-

erklärtem, teils systematischen *Rauschen*. Zur Integration dieser Befunde in einen einheitlichen theoretischen Rahmen werden nachfolgend alternative, handlungstheoretische Ansätze zur RCT vorgestellt, die einen integrativen Ansatz verfolgen.

Die RCT und alle ihre Varianten gehen von der Existenz einer zentralen Exekutive (Baddeley & Hitch, 1974) aus, die die Handlungsoption mit dem höchsten, subjektiv erwarteten Verstärkerwert selektiert und ausführt. Oben wurden RCT-orientierte Erklärungsansätze aufgeführt, die dieses theoretische RCT-Gerüst beibehalten, jedoch dessen Modellparameter erweitern und dadurch auch soziale Präferenzen, subjektive Erwartungen und periphere Informationsverarbeitungsprozesse erklären können. Jedoch existiert auch weiterhin Empirie, die der RCT und ihren Grundannahmen bezüglich menschlicher Entscheidungsprozesse und Handlungssteuerung grundlegend zu widersprechen scheint. Jenseits des RCT-Verhaltensparadigmas haben sich unter anderem in Soziologie, (Sozial-)Psychologie, Wahrnehmungspsychologie und Neurowissenschaften eine Vielzahl an (ebenfalls individualistischen, mikro-orientierten) Konzepten geringer und mittlerer Reichweite zu Aspekten der Informationsverarbeitung und Handlungssteuerung, aber auch übergreifende theoretische Rahmenmodelle etabliert, die sich grundlegend von den Annahmen des RCT-Universums unterscheiden. Ausgangspunkt vieler dieser Konzepte bilden die angenommenen Informationsverarbeitungsprozesse, die die Grundlage der Bewertung einer Situation sowie des Handelns darstellen. Anhand des psychologischen BRAC-Rahmenmodells (Binding and Retrieval in Action Control) (Frings et al., 2020a, 2020b) wird im Folgenden die Bedeutung alternativer Wahrnehmungs- und Informationsverarbeitungsprozesse für die Handlungssteuerung im ÖGS skizziert. Daran anschließend werden Dimensionen für die Klassifizierung von Konzepten vorgestellt und auf das soziologische Modell der Frame-Selektion (MFS) (Esser, 1996, 2001a, 2001b, 2002a, 2002b, 2003, 2010a, 2010b, 2018, 2019; Kroneberg, 2005, 2011, 2014; Esser & Kroneberg, 2015) projiziert, das bereits auf das ÖGS angewendet wurde. Dies leitet über zur Fragestellung, ob sich anhand dieser alternativen Konzepte möglicherweise einige der bisher nicht hinreichend erklärten Befunde zum ÖGS leistungsfähiger erklären lassen als durch das theoretische Konstrukt konditionaler Kooperation, das argumentativ auf der RCT und dem Optimierungskalkül basiert.

Funktion der Informationsaufnahme und des Gedächtnisses

Eine grundlegende Eigenschaft von Informationen aus der Umwelt ist, dass sie aus kleinen, aktuellen, teils inkonsistenten Ausschnitten bestehen. Beispielsweise werden akustische Sinneseindrücke der gleichen Quelle von beiden Ohren zeitlich versetzt aufgenommen. Visuelle Informationen bestehen aus Bruchstücken, die während der Fixationsdauer des Auges erfasst und über die parallelen Sehbahnen ins Gehirn geleitet werden (Kandel, Schwartz, Jessell, Siegelbaum, & Hudspeth, 2013). Die Funktion, die Informationsaufnahme erfüllen muss, ist die Widerspiegelung der Umwelt in unserem Bewusstsein als geschlossenes Abbild (Hoffmann, 2013). Das Ge-

gedächtnis ist hierbei keine gesonderte Leistung, sondern Teil der Prozesse zur Widerspiegelung von Umwelteindrücken im Bewusstsein. Es steht in engem Bezug zum Verhalten (Engelkamp & Cohen, 1991).

Die ältere, psychologische Gedächtnisforschung folgt passiven Speichertheorien (Gräsel & Gniewosz, 2015), die das Gedächtnis als die Fähigkeit definieren, aufgrund derer vergangene Informationen nach einer Behaltenszeit wieder reproduziert werden können. Viele Autoren der aktuellen Gedächtnisforschung betrachten das Gedächtnis demgegenüber nicht als passiven Speicher, sondern als aktive Tätigkeit. Sie dient dazu, durch Verfügbarmachung vergangener Eindrücke die aktuellen, zeitlich getrennten Ausschnitte (z. B. bei Fixationen des Auges) zu einem geschlossenen Abbild der Umwelt zu integrieren. Umgekehrt „erfolgt die Speicherung der Eindrücke integrativ in einem die objektiven Zusammenhänge widerspiegelnden Ganzen und nicht isoliert voneinander und einzeln“ (Hoffmann, 2013, S. 9). Der Abruf von Gedächtnisinhalten ist damit ein konstruktiver Prozess – mit jedem Aufruf dünnt eine Erinnerung emotional aus. Die Vorversion wird nach jedem Abruf anschließend neu überschrieben. Hierbei sind die Übergänge zwischen Erinnerung und Wahrnehmung fließend – dies wird bestätigt durch die Konvergenz der Wahrnehmungs- und Imagery-Forschung: Je anschaulicher und plastischer Ereignisse erinnert oder sich vorgestellt werden (und je suggestiver die Information, z. B. unter Hypnose ist), desto ähnlicher sind die Hirnareale, die an diesen Prozessen beteiligt sind (Kosslyn, Ganis, & Thompson, 2001).

Die Widerspiegelungstätigkeit besteht damit aus der Integration von internen und externen Reizen und abgerufenen Bewusstseinsselementen (Retrieval von Speicherinhalten) zu einem geschlossenen Abbild. Sowohl dieses integrierte Abbild aus Sinneseindrücken und kognitiven Prozessen als auch der Abruf gespeicherter Merkmale vergangener Reize, Reaktionen und Handlungskonsequenzen sowie deren Integration zu größeren Einheiten sind notwendig, um adaptiv handeln zu können. Konkret: um im experimentellen ÖGS sinnvoll handeln zu können, ist sowohl eine Integration der bruchstückhaften Informationen zu einer sinnvollen inneren Repräsentation der Spielsituation notwendig, als auch der Abruf der Merkmale von Reizen und Reaktionen sowie vergangener und zukünftig erwarteter Konsequenzen.

Während die Bezüge zwischen Handeln und isolierten Aspekten des kognitiven Systems (Aufmerksamkeit [Allport, 1987], Wahrnehmung [Gegenfurtner, Bremmer, Fiehler, Henriques, & Krauzlis, 2010; Schütz, Braun, & Gegenfurtner, 2011], Informationsverarbeitung [Fazio, 1990], Lernen und Gedächtnis [Kandel, 2007]) breit erforscht sind, rückt in jüngerer Zeit eine integrierte Betrachtung aller Aspekte des kognitiven Systems zunehmend in den Fokus. Ausgehend vom ideomotorischen Prinzip (James, 1890/2007), dass Handlungen „durch die Antizipation ihrer beabsichtigten perzeptuellen Effekte (der ‚Idee‘) kodiert, ausgewählt und gesteuert werden“ (Frings et al. 2020a, S. 2), geht der *Common-Coding-Ansatz* (Prinz, 1990; 1997) von einer Überlappung der internen Kodierung von Handlung und Wahrnehmung aus. Leitidee des *Common-Coding-Ansatzes* ist, dass Phänomene aus unterschiedlichen Aspekten kognitiver Prozesse (hier: wahrgenommene Ereignisse und geplan-

tes Handeln) eine gemeinsame Repräsentation im kognitiven System haben. Die Vorläufer dieser Forschung zur Merkmalsintegration stammen aus dem Bereich der visuellen Wahrnehmung (Hommel, 1999). Kahneman, Treisman und Gibbs (1992) postulieren dazu beispielsweise, dass die Wahrnehmung eines visuellen Objekts zum Anlegen einer sogenannten Objekt-Datei (object-file) führt. Es formiert sich so gesehen eine episodische Gedächtnisspur. Weitere Forschung (siehe z. B. Hommel, Lippelt, Gurbuz, & Pfister, 2017; Janczyk, 2016; Koch & Kunde, 2002; Kunde, 2006) postuliert hierauf aufbauend, dass im kognitiven System Merkmale von Reizen, Merkmale von Reaktionen und Merkmale der aus der Reaktion resultierenden Effekte in eine gemeinsame, komplette Handlungsepisode integriert werden. Die mentale Repräsentation einer Episode umfasst hierbei sowohl einfache sensorische Merkmale der Reize, Reaktionen und Effekte, als auch Merkmale im weiteren Sinne, z. B. affektive oder semantische Merkmale.

Episoden und Aspekte der Handlungskontrolle werden vom BRAC-Rahmenmodell (Binding and Retrieval in Action Control) (Frings et al., 2020a; 2020b) miteinander in Verbindung gesetzt. Das psychologische BRAC-Rahmenmodell beschreibt paradigmengreifend zentrale Prozesse der Informationsverarbeitung und der Handlungskontrolle und erhebt den Anspruch, eine Vielzahl von teils unverbundenen Phänomenen und Konzepten der Psychologie anhand einheitlicher Mechanismen zu erklären. Kernelemente des BRAC-Rahmenmodells sind der Abruf (Retrieval) und die Integration von Episoden sowie die Unterscheidung zwischen willkürlichen und automatischen Prozessen. Abbildung 1 zeigt am Beispiel des wiederholten ÖGS die Basis-Mechanismen des BRAC-Rahmenmodells.

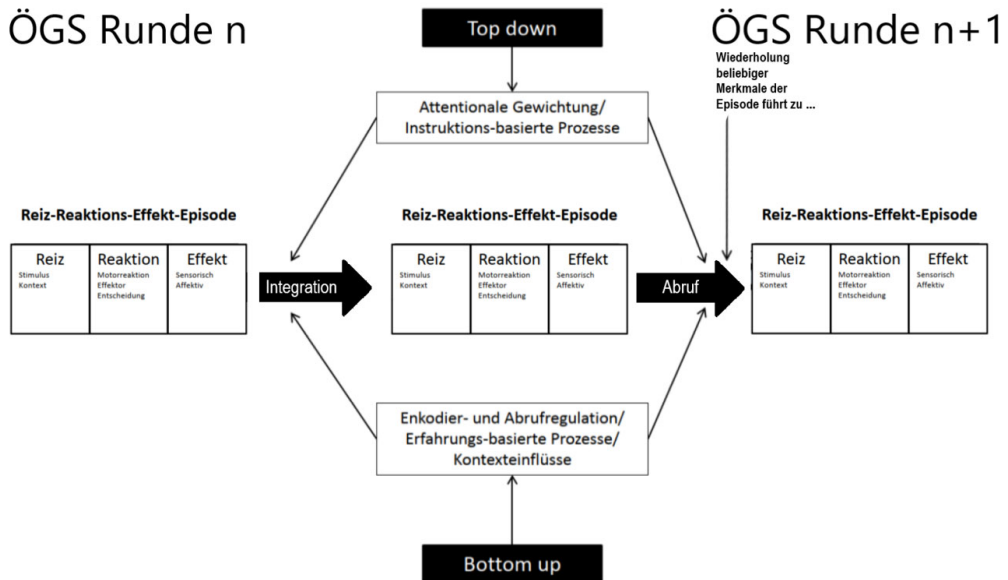


Abbildung 1 Schematische Darstellung des ÖGS im BRAC-Rahmenmodell (vgl. Frings et al. 2020a, S. 5-8)

Episoden im BRAC-Rahmenmodell bestehen aus Reiz, Reaktion und Effekt (Frings et al., 2020a, S.3). Im ÖGS wären dies beispielsweise der experimentelle Stimulus (Reiz), der vom sensomotorischen System ausgeführte Tastendruck und dessen taktiles Feedback (Reaktion) sowie die Konsequenzen der Entscheidung beispielsweise in Form des auf dem Computerbildschirm des Experimentallabors angezeigten Feedbacks (Effekt). Allen Teilen dieser im Bewusstsein präsenten Episode werden Merkmale zugeschrieben, beispielsweise für den Effekt die erwarteten Auszahlungen (Erwartungswerte in der RCT wären damit im BRAC-Rahmenmodell als Effekt-Merkmal codiert). Nachdem Runde n beendet ist, werden alle internen und externen Informationen in die Episode integriert und abgespeichert. Ist alles wie erwartet abgelaufen, festigt sich die ursprüngliche Episode. Der Start der Folgerunde $n+1$ (Stimulus) stellt nun ein Merkmal dar, das fest mit der Episode verbunden ist und dazu beiträgt, dass diese erneut abgerufen wird. Falls jedoch (i) das Feedback (Effekt in Runde n) eine (positive oder negative) Überraschung ergeben hat (Merkmal des Effektes), oder falls beispielsweise (ii) in Runde $n+1$ ein erwartetes Feedback ausgeblieben ist (Störreiz), besteht zwar eine Tendenz zur Änderung der Episode, das BRAC-Rahmenmodell prognostiziert jedoch eine starke Tendenz, Distraktoren (z. B. Störreize) unterzugewichten und die ursprüngliche Episode (trotz Unangemessenheit) fortzusetzen (weiter unten im Modell der Frame-Selektion als *Modellstärke* bezeichnet), solange keine sehr starken Merkmalen einer alternativen Episode vorliegen.

Das BRAC fokussiert bewusst auf die grundlegenden Mechanismen, um sie für spezifische Konzepte als übergreifendes Rahmenmodell anwendbar zu machen. Hierbei beschränkt sich das BRAC auf die Betrachtung von Episoden, ohne Aussagen über kleingranularere Bewusstseinsselemente zu machen. Auch werden die Mechanismen zum Abruf und zur Integration von Episoden und der mit ihnen verbundenen Merkmale nicht genauer spezifiziert. Jedoch geht das BRAC vor dem Hintergrund seiner teils behavioristischen Wurzeln von eher passiven Mechanismen des Abrufs und der Integration aus (Frings et al., 2020a). Um sinnvolle Aussagen über das Verhalten von Spielern im ÖGS ableiten zu können, werden nachfolgend weitere Konzepte skizziert, die kleingranularere Bewusstseinsselemente adressieren und Selektionsmechanismen konkretisieren. Das Modell der Frame-Selektion (Esser, 1996, 2001a, 2001b, 2002a, 2002b, 2003, 2010a, 2010b, 2018, 2019; Kroneberg, 2005, 2011, 2014; Esser & Kroneberg, 2015) geht hierbei stärker als das BRAC von aktiven Selektionsmechanismen aus, die auf eine Vielzahl von Typen kognitiver, konativer und affektiver Prozesse anwendbar sind, und hebt zudem die Trennung des BRAC bezüglich Bottom-up- und Top-down-Prozessen auf, indem Individuen aktiv entscheiden können, welche Informationsverarbeitungsprozesse automatisiert oder rational gesteuert erfolgen sollen.

Klassifizierung von Verhaltenstheorien

Verhaltenstheorien bestehen sowohl aus expliziten als auch aus impliziten Modellparametern und Annahmen, die sich entlang der folgenden Dimensionen klassifizieren lassen: (a) Aufnahme und Verarbeitung (z. B. Integration) von Reizen, (b) Typ von Bewusstseinsselementen (z. B. Frame, Einstellung, Motiv) und Granularität von Bewusstseinsselementen (z. B. Episode, Entscheidung, elementare Emotion) sowie (c) Mechanismen des Zusammenwirkens zwischen mentalen Prozessen und Verhalten.

Die RCT geht, wie oben dargestellt, implizit von einer zentralen Exekutive aus, die die Informationsverarbeitung, die Bildung von Erwartungswerten und die Auswahl und Ausführung der selektierten Handlungen steuert. Aus RCT-Perspektive existiert zwingend eine zentrale Exekutive, die die nutzenmaximale Handlungsoption berechnet und selektiert. Diese nutzenmaximale Entscheidungsalternative ist automatisch die Handlungsintention (Behavioral Intention BI) und löst den Willen zur Handlungsausführung (*urge*) aus. Die zentrale Exekutive übt zudem auch die Handlungskontrolle aus. Eine Vielzahl von Argumenten und umfangreiche empirische Evidenz sprechen gegen diese Argumentation. Ein Beispiel hierzu ist die Literatur zur Grundsatzfrage der Neurowissenschaften, ob das Bewusstseinssystem das Verhalten kausal steuert, also im verbundenen Modus (*connected mode*) willentlich die Handlungen ausführt. Nach Entdeckung des Bereitschaftspotentials (Kornhuber & Deecke, 1965) konnte jedoch eine inverse Kausalität von Entscheidungen und Verhalten robust nachgewiesen werden (Frith & Haggard, 2018; Haggard & Eimer, 1999; Libet, Gleason, Wright, & Pearl, 1983; Nann, Cohen, Deecke, & Soekadar, 2019). Dies hat in der Folgezeit in vielen Fachdisziplinen, einschließlich der Rechtswissenschaften, eine breite Debatte über Willensfreiheit und Handlungskontrolle ausgelöst. Doch auch wenn Verhalten nicht unmittelbar durch das Bewusstseinssystem gesteuert wird, besteht eine enge, interdependente Verbindung. Im Kontext der oben dargestellten Befunde kann hierbei zwischen zwei Arten von Handeln unterschieden werden: (1) Die Wahrnehmung vergangenen und die Erwartung zukünftigen eigenen Handelns im Bewusstsein und (2) die durch das Bereitschaftspotential eingeleitete und durch das sensomotorische System gesteuerte tatsächliche Handlungsausführung. Die Handlungssteuerung erfolgt demnach nicht durch eine zentrale Exekutive, die vergleichbar einem Puppenspieler die Handlungen steuert, sondern durch (1) den Aufruf von Episoden bzw. die Aktivierung gedanklicher Modelle, die sich (2) mit dem Handlungssystem synchronisieren (*connected mode*). Erfolgt keine Handlungskontrolle, ist dennoch Verhalten beobachtbar (*non-connected mode*) und die entsprechenden sensomotorischen Systeme können sich beispielsweise in einer Orientierungsfunktion oder im automatischen Modus befinden.

Der Match als alternativer Selektionsmechanismus mentaler Elemente

Ausgangspunkt für die folgende Betrachtung des Mechanismus, der bei der Selektion von Entscheidungen und Handlungen zum Tragen kommt, ist die Frage, ob sich Bewusstseinsselemente unterschiedlicher Ausprägung und unterschiedlicher Granularität grundsätzlich voneinander unterscheiden. Falls keine grundsätzlichen Unterschiede bestehen, erscheint eine getrennte Modellierung der Selektionsmechanismen künstlich – für alle mentalen Elemente sollten in diesem Fall die gleichen Mechanismen des Abrufs (Retrieval), der Integration und der Speicherung gelten.

Die RCT geht von einer inhaltlichen Zweiteilung der Bewusstseinsselemente aus: Auf der einen Seite stehen die exogenen, unabhängigen Bewusstseinsselemente, auf die die zentrale Exekutive (vergleichbar einem Homunkulus) wenig (Motive, Präferenzen) oder keinen (Wahrnehmung, Framing, Emotionen, Persönlichkeitsfaktoren) Einfluss hat. Diesen Bewusstseinsselementen stehen die von der zentralen Exekutive aktiv evaluierten und selektierten Elemente wie Erwartungen oder Nutzen gegenüber. Auch das BRAC-Rahmenmodell bezieht seine Mechanismen des Abrufs und der Integration nur auf einen Ausschnitt aller möglichen Bewusstseinsselemente, indem es seinen Untersuchungsgegenstand auf Episoden als Konstrukte auf der höchsten Granularitätsebene begrenzt.

Das soziologische Modell der Frame-Selektion⁴ (MFS) von Hartmut Esser (1996, 2001a, 2001b, 2002a, 2002b, 2003, 2010a, 2010b, 2018, 2019; Kroneberg, 2005, 2011, 2014; Esser & Kroneberg, 2015) hebt den Dualismus einer inhaltlichen Klassifizierung von Bewusstseinsselementen in Teilen explizit auf. Stocké (2002) hat hierzu entscheidende Beiträge geleistet, indem er Framing-Theorien unter anderem auch anhand des Grades klassifiziert, zu dem sie von einem aktiven Retrieval der *Logik der Situation* ausgehen. Diese Klassifikation von Stocké wird von Esser (2002a) in Form einer passiven (automatisch-spontanen) und einer aktiven (rational-kalkulierenden) Frame-Selektion modelliert. Beide Selektionsmechanismen werden von Kroneberg & Esser (2015) in einer späteren Version des MFS nicht nur auf die Frame-Selektion, sondern nun auch auf die Skript-Selektion und die Handlungs-Selektion erweitert. Eine Anwendung des Konzepts auf Bewusstseinsselemente größerer (z. B. Episoden) oder kleinerer (elementare Bewusstseinsselemente, z. B. Emotionen) Granularität ist jedoch bisher nicht explizit erfolgt.

Die Alleinstellung des MFS begründet sich in seiner Modellierung des *Match* als (1) universalem Selektionsmechanismus zum Abruf (Retrieval) (2) einer Vielzahl

4 Das MFS ist eine Dual-Process-Theorie und erhebt den Selbstanspruch einer allgemeinen Handlungstheorie (General Theory of Action), die einen zweistufigen Entscheidungsprozess modelliert. Auf beiden Stufen können sowohl zentrale als auch periphere Informationsverarbeitungsprozesse stattfinden. Das MFS behält einerseits Elemente aus der RCT bei (Nutzenbegriff, Anschlussfähigkeit an die Verhaltensökonomie etc.), andererseits werden zentrale Elemente der RCT ersetzt (Aufgabe des Maximierungsprinzips zugunsten des *Match* als grundlegend andersartigem Selektionsmechanismus) und um weitere Elemente ergänzt (explizite Einbeziehung von kulturellen Konzepten in Form gedanklicher Modelle; aktive Selektion beispielsweise von Wahrnehmungsschemata, Route der Informationsverarbeitung und [komplexen] Handlungsskripten).

unterschiedlicher kognitiver Prozesse, Schemata, mentaler Modelle und Bewusstseinselemente, beispielsweise Frames, Einstellungen, Entscheidungen, Handlungsskripte bis hin zur Wahl der passenden Route der Informationsverarbeitung. Die Kernfaktoren für den *Match* m und damit die Stärke der Aktivierung eines Schemas (z. B. Framing als Kooperationsspiel), gedanklichen Modells (z. B. passendes Gerechtigkeitsprinzip) oder Handlungsskripts (z. B. Beitragshöhe zum gemeinsamen Projekt) sind die interne Zugänglichkeit a des gedanklichen Modells aus der Intensität der Encodierung, die Salienz e dieses mentalen Modells in einer spezifischen Situation durch die Anwesenheit bestimmter Symbole sowie das Fehlen störender Stimuli u (Esser, 2001b, S. 270). Ergibt sich bei der multiplikativen Verbindung dieser drei Faktoren $m = 1$, spricht Esser von einem perfekten *Match* des betreffenden gedanklichen Modells mit der Situation, der dann zur Auslösung des passenden Frames führt und hiernach zum entsprechenden Verhalten. Für den perfekten *Match* (im automatisch-spontanen Modus) gilt als wichtigste Implikation „die Überschreibung aller anderen Präferenzen und Erwartungen bzw. die Aktivierung einer ggf. ganz anderen Nutzenfunktion bei einem perfekten *Match* – bis hin zur vollständigen *Unbedingtheit* der Imperative für den Frame der betreffenden Situation“ (Esser, 2018, S. 14). Das MFS sieht andererseits die Integration der orthodoxen RCT vor. In diesem Zweig werden (begrenzt) rational im reflexiv-kalkulierenden Modus (rc-Modus) zukünftige Konsequenzen einer Handlung abgewogen. Die Abwägung im rc-Modus kann durch eine starke Motivation hervorgerufen werden (z. B. „Hier geht es um viel Geld“) oder durch einen *Mis-Match* der Situation (z. B. kein Mitspielender trägt zum gemeinsamen Projekt bei) mit dem bisher bestehenden gedanklichen Modell (z. B. sich in einer kooperativen Spielsituation zu befinden). Unter diesen Bedingungen kann es entsprechend der Annahmen des MFS zu einem *Reframing* kommen. Dazu muss der aktuelle Frame stark genug gestört sein, und es muss vor allem ein ausreichend attraktives alternatives gedankliches Modell zur Verfügung stehen. Die Abwägung findet hierbei entsprechend den Annahmen von Wert-Erwartungs-Theorien statt.

Für das soziologische Modell der Frame-Selektion (MFS) ist die Annahme einer zentralen Exekutive nicht zwingend notwendig. Der *Match* als grundlegender Selektionsmechanismus stellt kein Optimierungskalkül, sondern die Passung eines gedanklichen Modells an die subjektive Definition der Situation dar. Der *Match* ist hierbei analog auf das BRAC-Modell anwendbar und kann als Mechanismus zur Integration und zum Abruf von Episoden interpretiert werden. Bezogen auf das ÖGS stellt das MFS ein Rahmenmodell dar, in das die RCT, aber auch alternative mentale Modelle integriert werden können. Angewendet auf das individuelle Entscheidungsverhalten im ÖGS bestehen aus Perspektive des MFS keine modelltheoretischen Restriktionen, wie Spieler eine Situation bewerten und sich entsprechend ihrer subjektiven Wahrnehmung der Situation entscheiden.

Forschungsfragen und Hypothesen

Wie oben dargestellt, wurde lange Zeit in der Literatur davon ausgegangen, dass das experimentelle ÖGS weitgehend erforscht sei, da sich seine Ergebnisse auch gegenüber Designvarianten als robust erwiesen hatten und konditionale Kooperation als wichtigster Erklärungsmechanismus vielfach bestätigt werden konnte. Bei einer Sekundäranalyse der Rohdaten vieler ÖGS-Experimente zeigen sich auf Individualebene jedoch Entscheidungsmuster und Einstellungen, die nicht mit der RCT im Einklang stehen. Zunächst wurde daher in der obigen Einführung die Frage nach den Grenzen und der Leistungsfähigkeit des experimentellen ÖGS zur Messung von konditionaler Kooperation diskutiert, um methodeninduzierte Fehlerquellen auszuschließen. Anschließend wurden handlungstheoretische Rahmenmodelle aus Psychologie und Soziologie exemplarisch vorgestellt, die die RCT umfassen und somit möglicherweise das ÖGS leistungsfähiger erklären können. Diese Rahmenmodelle gehen davon aus, dass Individuen aktiv die Rahmung, die Episode und die situativen Persönlichkeitszüge bzw. den Verhaltenstyp aufrufen, die situativ passen. Typen, z. B. *konditionale Kooperationsure*, sind aus dieser Perspektive nicht exogen vorgegebene Persönlichkeitsmerkmale, sondern Bündel an heterogenen Verhaltensmustern, hinter denen sich jeweils sehr unterschiedliche Motive und gedankliche Modelle verbergen können. Konkret, über die subjektive Logik der Situation (z. B. Wahrnehmung von Feedback) und die passende Episode bzw. das mentale Modell (z. B. konditionale Kooperation) entscheiden nicht die Reize (objektive Spielsituation, Geldauszahlungen oder gegebene Feedback-Informationen), sondern der Spieler (auf einer zentralen oder peripheren Route der Informationsverarbeitung). Ein Spieler kann daher subjektiv eine konditional kooperative Strategie verfolgen, auch wenn dies objektiv nicht möglich ist, weil kein Feedback gegeben wird. Zudem schränken Rahmenmodelle weder die Ausprägungsform noch die Anzahl möglicher Typen, Verhaltensmuster und mentaler Modelle ein, die beispielsweise durch Sozialisation entstehen können. Rahmenmodelle prognostizieren *ex ante*, dass äußere und innere Störreize, z. B. gruppenspezifisches Feedback, der erneute Start eines Experimentes, eine sehr häufige Wiederholung oder das Ausbleiben eines erwarteten Feedbacks sowohl individuell als auch kollektiv zu einem Austausch der aktivierten Typen, Episoden und gedanklichen Modelle führen können. Theoriegeleitet können Rahmenmodelle daher potentielle Schwachstellen des ÖGS *ex ante* identifizieren, bei denen beispielsweise ein Umschlagen der *Logik der Situation* zu erwarten ist.

Die grundlegenden Forschungsfragen der vorliegenden Untersuchung sind:

- (1) Stehen in Abhängigkeit von Person und Situation hinter dem gleichen Verhaltenstyp unterschiedliche Motive und Intentionen? Korrespondiert der Verhaltenstyp mit den gemessenen Verhaltensgründen (Einstellung = Verhalten)?
- (2) Stellt das wiederholte ÖGS ein Treatment dar oder handelt es sich je nach gruppenspezifischem Verlauf um völlig unterschiedliche Treatments mit jeweils eigener Gruppenkultur, individuellen *Erlebniswelten* und unterschiedlichen Sets und

Anteilen aktivierter Verhaltensstrategie-Typen? Konkret, wie denkt und handelt ein *unkonditionaler Kooperationspartner*, wenn er in eine Gruppenkultur der konditionalen Kooperation gezwungen wird? Falls sich ein *unkonditionaler Kooperationspartner* an die Gruppenkultur anpasst – bleibt seine Identität dann weiterhin *unkonditionaler Kooperationspartner*, weil er sich anfänglich entsprechend verhalten hat, oder ist vielmehr Essers (2017, S. 8) Konzept situativer Persönlichkeitszüge treffender und erklärungsstärker?

- (3) In der Literatur wird meist behauptet, dass eine Mehrheit der Spieler im ÖGS dem Typ *konditionaler Kooperationspartner* zuzuordnen ist und lediglich in der ersten Runde gemäß ihrer *home-grown beliefs* entscheidet. Aus Rahmenmodellen lässt sich ableiten, dass Individuen nicht die objektiv, sondern die subjektiv zur Spielsituation passenden Episoden, Schemata, mentalen Modelle und Verhaltenstypen (z. B. konditionale Kooperation) abrufen. Daraus resultiert die Forschungsfrage, welche Entscheidungsprozesse in einem Treatment ganz ohne jegliches Feedback ablaufen, in dem Spieler des Typs *konditionaler Kooperationspartner* objektiv nicht konditional kooperieren können.

Hypothesen

Die RCT definiert *konditionale Kooperationspartner* als situationsübergreifenden Typ, der sich aus einer Mischung prosozialer und eigennutzmaximierender Präferenzen ergibt. Auch wenn im ÖGS sehr viele Runden gespielt oder das Feedback variiert wird, prognostiziert die RCT, dass der Typ konstant bleibt und sich lediglich die Entscheidungen anpassen. Rahmentheorien prognostizieren demgegenüber, dass Individuen ihre mentalen Modelle, Präferenzen, Motive und Strategien und auch ihren Typ ändern, wenn alternative Modelle subjektiv besser passen (*matchen*). Nachfolgend werden für Spielertypen, Verhaltensmuster und Verhaltensgründe Hypothesen abgeleitet.

Typen-Hypothese

Autoren in der Tradition von FGF (2001) gehen davon aus, dass die abhängige Variable (AV) Typ durch die unabhängige Variable (UV) ÖGS aktiviert wird. Typen stellen aus dieser Perspektive ein Persönlichkeitsmerkmal dar, das den individuellen Präferenzen entspricht und beispielsweise in Kooperationsspielen (passiv) aktiviert wird. In der ersten Runde verhalten sich die Vpn entsprechend den Annahmen von FGF (2001) nach ihren *home-grown beliefs*, und es bestehen daher keine Unterschiede zwischen den Treatments oder zwischen den Gruppen. Die Typen sind in allen Treatments und Gruppen gleichverteilt. Folgende Hypothese leitet sich hieraus ab:

H1a: In der ersten Runde besteht zwischen den Treatments kein signifikanter Unterschied im Beitragsverhalten der Vpn.

Aus den oben in den methodischen Überlegungen formulierten Annahmen Essers (2018) zum subjektiven Framing der Spielsituation ergibt sich für die Anfangsrunde:

H1b: Im ÖGS sind in der ersten Runde die Anteile von Altruisten mit einem Altruismus-Frame (AF-Typ) und Egoisten mit einem Egoismus-Frame (EF-Typ) gleichverteilt.

Aus Perspektive von Rahmentheorien wählt ein Spieler den situativ zum Spieldesign und Treatment passenden Typ (Ausprägung der AV); jede Gruppe stellt aufgrund des gruppenspezifischen Feedbacks ab der zweiten Runde ein eigenes Treatment dar. In jedem Treatment existieren spezifische, unterschiedliche Typen. Hieraus leitet sich die nachfolgende Hypothese ab:

H1c: In Abhängigkeit von der Spielsituation (Treatment) und dem Spielverlauf (gruppenspezifische Kultur) variieren ab der zweiten Runde die Anteile der Typen.

Reframing-Hypothese

Jeder Spieler aktiviert im Sinne der Argumentation von Rahmentheorien zu Beginn des ÖGS die subjektiv passenden Bewusstseinsselemente und die jeweils korrespondierenden Verhaltensmuster und behält diese unverändert bei, solange sie der Situation entsprechen. Treten externe oder interne Stimuli bzw. Signale auf, die nicht zur ursprünglichen Wahrnehmung und Bewertung der Spielsituation passen, kann eine Neubewertung und gegebenenfalls ein *Reframing* des gedanklichen Modells erfolgen, das sich im Abruf der dann passenden Bewusstseinsselemente manifestiert. Bewusstseinsselemente umfassen hierbei mentale Modelle, Episoden, Frames, Emotionen, situative Persönlichkeitszüge, Einstellungen, aber auch die im Bewusstsein repräsentierten Verhaltensskripte sowie die passende Route der Informationsverarbeitung. Hieraus leitet sich die folgende, zweite Hypothese ab:

H2: Wenn die aktivierten Bewusstseinsselemente nicht mehr zur Spielsituation passen und gleichzeitig alternative Bewusstseinsselemente zugänglich sind, die attraktiver sind (*Match*), kommt es entsprechend der Annahmen des MFS zu einem *Reframing* der Bewusstseinsselemente und der damit verbundenen Verhaltensskripte.

Grad der RCT-Kompatibilität-Hypothese

Empirisch verhalten sich Individuen in feststehenden, vorgegebenen Bewertungsrahmen (z. B. im Wirtschaftsleben) zu einem hohen Grad entsprechend den RCT-Prognosen. Zudem besteht sowohl bei Spielern als auch bei Experimentatoren die Tendenz, das gezeigte Verhalten ex post als rational zu etikettieren. Dieses schon von Max Weber vorgebrachte Argument der Ex-post-Rationalisierung sozialer Prozesse (Münch, 1980) erschwert die Falsifizierbarkeit der RCT. Dennoch lassen sich aus der RCT und ihren Grundannahmen der Existenz einer zentralen Exekutive, die einem

Optimierungskalkül folgt, bei dem Erwartungswerte und Verhalten miteinander korrespondieren, grundlegende Kriterien aufstellen, die nicht verletzt sein dürfen. Dies umfasst die Stabilität und Konsistenz von Typen (z. B. *konditionale Kooperateure*), Entscheidungsmustern im Zeitverlauf (z. B. kein Trittbrettfahrerverhalten, das ohne Anlass zu UCH wechselt) und Verhaltensgründen sowie den Ausschluss theorieferner Typen, die nicht auf (eigennützigen, sozialen oder antisozialen) Präferenzen bzw. Gerechtigkeitsprinzipien basieren (z. B. heuristikbasierter Typ mit Beiträgen entsprechend der Glückszahl im Tageshoroskop). Für jedes Individuum lässt sich damit identifizieren, welche und wie viele dieser Kriterien zu welchem Schweregrad verletzt werden, und somit als Gesamtschau die Leistungsfähigkeit der RCT für das ÖGS abschätzen. Hieraus lässt sich folgende dritte Hypothese formulieren:

H3: Je mehr Runden im ÖGS gespielt werden und je genauer das Entscheidungsverhalten und die Entscheidungsgründe eines Individuums betrachtet werden, desto stärker werden grundlegende Annahmen der RCT verletzt.

Methode und experimentelles Design

Zur Überprüfung der Hypothesen wurden zwei Varianten des ÖGS gespielt. In beiden Treatments trifft jeder Spieler in einer Vierergruppe in jeder Runde die Entscheidung, welchen Anteil seiner Ausstattung mit 20 Punkten pro Runde er für sich behalten oder in das gemeinsame Projekt im ÖGS investieren will. Hierbei ist jeder Punkt 0.025 EUR (= 2.5 Cent) wert. Alle Beiträge, die in der jeweiligen Runde in das gemeinsame Projekt investiert werden, werden mit 0.4 multipliziert. Von dieser Summe erhält jeder Spieler ein Viertel, unabhängig davon, ob bzw. wieviel er zum gemeinsamen Projekt beigetragen hat.

Um die erste Hypothese (H1) zur Situationsabhängigkeit der Typen zu testen, wurde das wiederholte ÖGS ohne Bestrafung gespielt, nachfolgend in der Konnotation von Esser (2018) bezeichnet als NoP (No-Punishment)⁵, da es sich in den ersten 10 Runden (erstes Experiment, nachfolgend NoP-Exp1 genannt) um eine exakte Replikation des entsprechenden NoP-Treatments von F&G (2000) handelt. Für die Experimentdurchführung wurde die originale zTree-Programmierung (Fischbacher, 2007) genutzt. Die Prozeduren während der Durchführung der Sessions wurden konstant gehalten. Zudem wurden die Original-Fragebögen in Original-Typografie und Design benutzt, indem zunächst der schweizerische und der deutsche Fragebogen miteinander verglichen wurden, die dann um interne Inkonsistenzen und Übertragungsfehler bereinigt und an die Laborspezifika des Essener Labors für Experimentelle Wirtschaftsforschung (elfe) angepasst wurden⁶. Als Erweiterung

5 Die Eins-zu-eins-Replikation im NoP-Treatment in Experiment 1 (Runde 1-10) ermöglicht zudem einen unmittelbaren Vergleich mit den Ergebnissen von F&G (2000).

6 Beispielsweise die Beachtung aktueller ethischer Standards bezüglich der möglichst vollständigen und wahrheitsgemäßen Information der Vpn.

des Ursprungsdesigns wurde nach Abschluss von 10 Runden das NoP-Treatment erneut identisch über weitere 10 Runden (Experiment 2, nachfolgend NoP-Exp2 genannt) sowie danach nochmals über 10 Runden (Experiment 3, nachfolgend NoP-Exp3 genannt) gespielt. Im NoP-Treatment wurden den Spielern nach jeder Runde (einschließlich nach Abschluss jedes Experimentes) Informationen über den Gruppenbeitrag zum ÖGS, über die jeweiligen eigenen Gesamteinkünfte sowie über die Erträge nach Abschluss jedes Experimentes gegeben.

Im NoF-Treatment (No-Feedback) wurde bis zum Abschluss des zweiten Experimentes (Runde 20) kein Feedback gegeben (die ersten 10 Runden werden nachfolgend bezeichnet als NoF-Exp1, Runden 11 bis 20 nachfolgend als NoF-Exp2), ansonsten bestanden keine Unterschiede zum NoP-Treatment. Im NoF-Treatment gab es damit anders als im NoP-Treatment (bis Runde 20) weder ein Feedback über die Beiträge der anderen Gruppenmitglieder, noch über die kollektiven Gruppenbeiträge, noch über die jeweiligen Rundeneinkünfte der einzelnen Spieler, die sich aus der einbehaltenen Ausstattung sowie der Auszahlung an alle Gruppenmitglieder zusammensetzt. Im NoF-Treatment wurde den Teilnehmern nach der 20. Runde erstmals ein Feedback über die Beiträge der Gruppenmitglieder gegeben. Anschließend wurde das NoF-Treatment unverändert erneut über weitere 10 Runden ohne Feedback zwischen den Runden gespielt (Experiment 3, nachfolgend NoF-Exp3), sodass jeder Versuchsteilnehmer insgesamt 30 Runden gespielt hat. Beide Treatments bestanden somit aus drei unmittelbar hintereinander gespielten Experimenten mit jeweils 10 Runden, also NoP-NoP-NoP und NoF-NoF-NoF. Beide Treatments wurden im Partner Design gespielt.

Operationalisierung der Hypothesen

Um H1 zu testen werden die Vpn anhand der Höhe ihrer in der ersten Runde geleisteten Beiträge gemäß der Klassifikation von Esser (2018) in Egoisten mit einem Egoismus-Frame (EF-Typ) und Altruisten mit einem Altruismus-Frame (AF-Typ) eingeteilt. Dann werden die Vpn für die Runden 2 bis 30 entsprechend der aus den Verhaltenssequenzen abgeleiteten Typen von FGF (2001) sowie Thöni und Volk (2018) und Fallucchi et al. (2019) klassifiziert.

Um H2 zu testen, werden in Verschränkung zu den Ergebnissen zu H1 die Verhaltensgründe der Vpn erhoben. Im Anschluss an die insgesamt 30 Spielrunden werden anhand offener, qualitativer Fragen und quantitativer Skalen die Entscheidungsgründe und die subjektive Wahrnehmung eines Wechsels der subjektiven Wahrnehmung der Spielsituation bezogen auf die erste Runde und die 29 folgenden Runden abgefragt. Die Messinstrumente (siehe Anhang) wurden von den Autoren des Datensatzes (Betz, Esser, & Biniossek, 2020) aufbauend auf frühere eigene Studien entwickelt. Die quantitativen, bipolaren Items vom Likert-Type werden auf einer 7-stufigen Ratingskala von 1 = *stimme überhaupt nicht zu* bis 7 = *stimme voll und ganz zu* erhoben. Zur Erhebung der Prognosen der RCT dient das Item *Outcome-Maximierung*, zur Erhebung sozialer Präferenzen im Rahmen einer weiten RCT das Item

Fairness. Konditionale Kooperation analog zu FGF (2001) wird anhand der Items *Orientierung an den Mitspielenden* sowie *Reziprozität* gemessen. Zur Überprüfung von Rahmentheorien dienen die Messung der Stärke gedanklicher Modelle anhand des Items *Entscheidungsunsicherheit* sowie die Messung von *Emotionalisierung* als Hinweis darauf, ob die Situation mit dem gedanklichen Modell *matcht*.

Um H3 zu testen, wird im NoF-Treatment überprüft, ob steigende Beitragsverläufe identifizierbar sind, die gemäß der RCT nicht auftreten dürften. Für die Runden 2 bis 20 wird bis zum ersten Feedback geprüft, ob (A1) in mindestens einer der vorausgehenden Runden die Beiträge niedriger waren, (A2) die Beiträge über der unmittelbar vorangehenden Runde liegen, (A3) dieser Anstieg nach einem vorausgehenden Sinken erfolgt ist. Diese drei Kriterien haben den Charakter einer Guttman-Skala, da die Erfüllung des dritten Kriteriums die Erfüllung der beiden vorausgehenden mit einschließt. Einen besonders schweren Verstoß gegen die Annahmen der RCT stellen Beiträge dar, die höher als der Erstrundenbeitrag sind. Falls Beiträge oberhalb der Erstrundenbeiträge liegen, werden diese als (B1) kategorisiert. (B2) bezeichnet Beiträge, die sowohl höher als in der ersten Runde sind als auch oberhalb der unmittelbar vorangehenden Runde liegen. Ebenfalls der Guttman-Skala-Logik folgend ergibt sich schließlich (B3) für gestiegene Beiträge nach unmittelbar zuvor erfolgtem Sinken, die zusätzlich oberhalb des Erstrundenbeitrags liegen. Für NoF-Exp3 wird für die Runden 22 bis 30 das gleiche Kategorienschema analog genutzt, weil in Runde 21 zumindest aus Sicht konditionaler Kooperation eine Anpassung nach oben zulässig ist und erst ab Runde 22 bei ausbleibendem Feedback eine Erhöhung der Beiträge im Vergleich zu Runde 21 einen Verstoß gegen die RCT darstellt. Die Ergebnisse werden dann zu den von den Befragten angegebenen Verhaltensgründen in Beziehung gesetzt, um weitere Abweichungen von den RCT-Prognosen explorativ zu entdecken.

Durchführung

Die beiden experimentellen Treatments mit jeweils 16 Versuchspersonen (Vpn) ($N = 32$) wurden in zwei Sessions (4. und 8. Juli 2019) im elfe Labor für Experimentelle Wirtschaftsforschung an der Universität Duisburg-Essen erhoben. In den beiden Sessions wurden gleichzeitig beide Treatments mit jeweils zwei Gruppen zu je vier Vpn gespielt. Die Durchführung erfolgte im Rahmen einer größeren Replikationsstudie (Betz et al., 2020). Die Vpn (einschließlich der drei Ersatzspieler pro Session) wurden aus dem Versuchsteilnehmerpool des elfe Laboratoriums mit der Rekrutierungssoftware ORSEE (Greiner, 2015) rekrutiert. Bei der Rekrutierung wurde auf eine Gleichverteilung der Geschlechter sowie die erstmalige Teilnahme an einem ÖGS geachtet. Beim Eintritt in das Labor wurden die Vpn ihren Computerplätzen zugewiesen (Ziehung nummerierter Tischtennisbälle) bzw. bei Auslosung als Ersatzspieler mit 10 EUR ausbezahlt. An jedem der isolierten Computerplätze lag analog zum Experimentaldesign von F&G (2000) ein Instruktionsbogen einschließlich Kontrollfragen aus, die mit Kugelschreiber beantwortet werden mussten. Sowohl das Ex-

periment als auch die Befragung (Questionnaire) nach Abschluss des Experimentes wurden mit der Software z-Tree (Fischbacher, 2007) durchgeführt.

Ergebnisse

Deskriptive Ergebnisse

Das Durchschnittsalter der zehn männlichen und sechs weiblichen Teilnehmer im NoP-Treatment beträgt 25.25 Jahre. Im NoF-Treatment beträgt das Durchschnittsalter der acht männlichen und acht weiblichen Vpn 25.06 Jahre. Unter den Teilnehmern befinden sich ausschließlich Studenten aus dem Versuchsteilnehmerpool des elfe Laboratoriums der Universität Duisburg-Essen. Im NoP-Treatment beträgt die durchschnittliche Fachsemesterzahl 6.31. Im NoF-Treatment ist die durchschnittliche Fachsemesterzahl 8.5. Im NoP-Treatment studieren sieben Vpn (43.75 %) an der Fakultät für Wirtschaftswissenschaften, vier Vpn (25 %) studieren auf Lehramt, die übrigen fünf Vpn stammen aus den Fachbereichen Biologie, Psychologie, Medizin und Geisteswissenschaften. Im NoF-Treatment studieren fünf Vpn (31.25 %) an der Fakultät für Wirtschaftswissenschaften, vier Vpn (25 %) studieren auf Lehramt, jeweils zwei Vpn (12.5 %) studieren Biologie bzw. Medizin und jeweils eine Vpn Mathematik bzw. Innopreneurship. Nach Abschluss der etwa 45 Minuten dauernden Experimente wurden die Nummern der Computerplätze einzeln aufgerufen und die Spieler in bar ausbezahlt, die Auszahlung betrug für das NoP-Treatment durchschnittlich 21.51 EUR mit 15.20 EUR Minimum und 28.20 EUR Maximum sowie für das NoF-Treatment durchschnittlich 17.78 EUR (Minimum 9.90 EUR und Maximum 24.00 EUR). Die Daten wurden mithilfe der Software SPSS ausgewertet und angelehnt an die Method of Difference (Lieberson, 1994; Lijphart, 1971; Mill, 1843) analysiert.

In Abbildung 2 sind die aggregierten Beiträge zum gemeinsamen Projekt im ÖGS für die beiden Treatments dargestellt. Die Aggregation über Gruppengrenzen hinweg ist für das NoP-Treatment (und ab Runde 21 auch für das NoF-Treatment) methodisch nicht zulässig, da die Gruppen, wie oben gezeigt, aufgrund ihrer strukturell verschiedenen Beitragsverläufe unterschiedliche Treatments darstellen; dennoch werden hier die aggregierten Beiträge präsentiert, um den Vergleich zur in der Literatur üblichen visuellen Darstellung auf Treatment-Ebene zu ermöglichen. In Abbildung 2 sind in beiden Treatments relativ hohe Anfangsbeiträge zu beobachten. Im NoP-Treatment liegen die Beiträge zu Beginn jedes Experimentes höher als am Ende, um im nachfolgenden Experiment erneut auf ein höheres Niveau zu springen – in der Literatur als isolierter Restart-Effekt bezeichnet (Chaudhuri, 2018; Esser, 2018). Im NoF-Treatment ist nach Ausbleiben von Feedback ab Runde 2 eine Art vorweggenommener Endrundeneffekt zu beobachten, sowie anschließend um dieses Niveau schwankende Beiträge. Während zwischen der ersten und der letzten Runde in Experiment 1 und 2 des NoF-Treatments (NoF-Exp1 und NoF-Exp2) ein fallender Trend besteht, ist in Experiment 3 (NoF-Exp3) ein steigender Trend

identifizierbar. Dabei gleichen sich die Beitragsniveaus von F&G (2000) und NoF, wohingegen NoP auf einem höheren Niveau liegt.

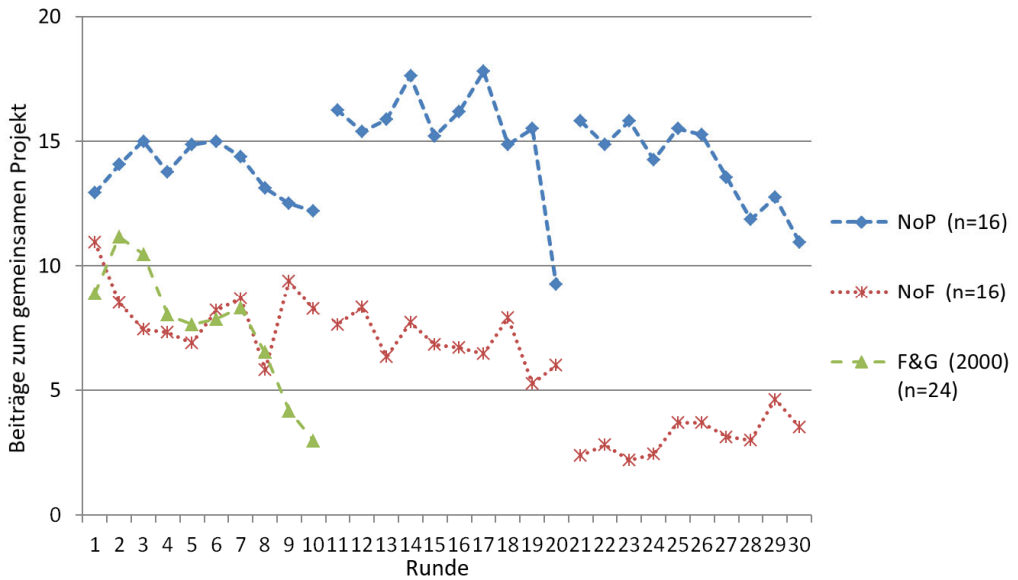


Abbildung 2 Durchschnittliche Beiträge zum gemeinsamen Projekt

Für alle sechs Experimente (NoP-Exp1, NoP-Exp2, NoP-Exp3, sowie NoF-Exp1, NoF-Exp2, NoF-Exp3) ergibt der Kolmogorov-Smirnov-Test für eine Stichprobe, dass die Nullhypothese einer Normalverteilung der Beiträge mit $p = .000$ höchst signifikant abzulehnen ist. Dies deckt sich mit der Sekundäranalyse der Daten von F&G (2000), die ebenfalls gemäß Kolmogorov-Smirnov-Test in den ersten 10 Runden höchst signifikant nicht normalverteilt sind ($p = .000$).

Des Weiteren bestätigt der Mann-Whitney-U-Test für NoP (Median = 20) und F&G (2000) (Median = 6) für die Runden 2 bis 10, dass die beiden Stichproben hochsignifikant unterschiedlich sind ($z = -7.438$, $p = .000$). Die Effektstärke nach Cohen (1992) liegt bei $r = .39$ und entspricht einem mittleren bis starken Effekt. Auch zwischen NoP und NoF (Median = 5) bestehen im Mittelwertvergleich für die Runden 2 bis 10 hochsignifikante Unterschiede ($z = -5.691$, $p = .000$). Die Effektstärke nach Cohen (1992) liegt bei $r = .36$ und entspricht einem mittleren bis starken Effekt. Dagegen zeigt sich zwischen NoF und F&G (2000) mittels Mann-Whitney-U-Test kein signifikanter Unterschied.

Überprüfung von Hypothese 1

Hypothese 1a: Erstrundenbeiträge (NoP & NoF)

In der ersten Runde treffen sowohl im NoF- als auch im NoP-Treatment alle Spieler ihre Entscheidung ohne vorheriges Feedback. Von ihrer Erstausrüstung von 20 Punkten tragen die Spieler in der ersten Runde des NoP-(NoF-)Treatments durchschnittlich 12.94 (10.94) Punkte zum Gemeinschaftsprojekt bei. Das NoF-Treatment entspricht damit den Erstrundenbeiträgen in anderen Studien von durchschnittlich 40 % bis 60 % auf Aggregatebene (Glassmann & Sauermann, 2011), die Erstrundenbeiträge im NoP-Treatment liegen geringfügig oberhalb dieser Spannweite. In Abbildung 2 wird jedoch deutlich, dass sowohl im NoP-Treatment als auch in NoF-Treatment in der ersten Runde höhere Beiträge geleistet werden als die durchschnittlich 8.88 Punkte bei F&G (2000). Diese höheren Beiträge resultieren aus den relativ hohen Beiträgen der männlichen Spieler. Männliche Vpn tragen im NoP-(NoF-)Treatment in der ersten Runde 15.2 (16.6) Punkte und weibliche Vpn 9.2 (5.3) Punkte bei⁷. In beiden Treatments (NoP und NoF) zusammengenommen tragen Männer in der ersten Runde signifikant ($z = -2.659$, $p = .008$) mehr zum gemeinsamen Projekt bei als Frauen.

Da sich das NoF-Treatment in der ersten Runde im Vergleich zum NoP-Treatment und zu F&G (2000) ausschließlich durch das Weglassen einer Textpassage im jeweiligen Instruktionsbogen unterscheiden, erscheint die gemeinsame Betrachtung der ersten Runde zulässig. In der ersten Runde des ersten Experimentes sollte sich das Verhalten zwischen den drei Treatments (F&G (2000), NoP und NoF) nicht signifikant unterscheiden (H1a), da erst ab der zweiten Runde die Unterschiede zwischen den Treatments evident werden. In einem Kruskal-Wallis-Test zeigt sich, dass zwischen den drei Stichproben in der ersten Runde statistisch kein signifikanter Unterschied besteht ($p = .228$). Insgesamt trifft H1a damit zu.

Hypothese 1b: Anteile EF-Typ und AF-Typ in erster Runde (NoP & NoF)

Bezüglich des Beitragsverhaltens in der ersten Runde finden sich im NoP-Treatment neun Vpn, also 56.25 %, die 20 Punkte zum gemeinschaftlichen Projekt beitragen, und eine Vpn (6.25 %), die zwischen 10 und 19 Punkte beiträgt. Damit sind akkumuliert 62.5 % der Spieler in der Kategorisierung von Esser (2018) dem Persönlichkeitstyp Altruist mit Altruismus-Frame (AF-Typ) zuordenbar. Die restlichen 37.5 % der Spieler sind als Egoisten mit Egoismus-Frame (EF-Typ) klassifizierbar, von denen drei Vpn (18.75 %) zwischen 1 und 9 Punkte und drei Vpn (18.75 %) null Punkte zum

7 Im NoF-Treatment, in dem nur vor der 21. Runde ein möglicherweise interagierendes Feedback über das Verhalten der Mitspieler gegeben wird, leisten männliche Vpn im Vergleich zu weiblichen Vpn in der ersten Runde einen 3.2-fachen Beitrag und über alle 30 Runden durchschnittlich einen 3.4-fachen Beitrag.

gemeinsamen Projekt beitragen. **H1b** trifft für die beiden erhobenen Treatments nicht zu.

Hypothese 1c: Gruppenkultur und gruppenspezifische Typen (NoP)

Während sich die Treatments zunächst in den Erstrundenbeiträgen nicht signifikant unterscheiden, erhalten die Spieler ab der zweiten Runde als gruppenspezifisches Treatment ein sehr unterschiedliches Feedback bezüglich der Beiträge ihrer Mitspielenden zum gemeinsamen Projekt im ÖGS und die Höhe der Beiträge nimmt anschließend je nach Treatment einen unterschiedlichen Verlauf. Hieraus folgt, dass die Mittelwertunterschiede somit nicht nur aus den Stichprobenunterschieden, Erstrunden-Entscheidungen oder AF-Typ bzw. EF-Typ resultieren, sondern (a) aus der Verteilung der Individuen auf die vier Gruppen des NoP-Treatments, (b) der Variation des Feedbacks und (c) dessen Wirkung auf die Individuen und ihr Beitragsverhalten. Das Beitragsverhalten führt letztendlich zu der entsprechenden (d) Einordnung in Verhaltenstypen entsprechend F&G (2000) bzw. Thöni und Volk (2018) bzw. Fallucchi et al. (2019). In welchem Ausmaß das Feedback im NoP-Treatment zwischen den Gruppen variiert, wird in Abbildung 3 gezeigt.

Für zwei der vier Gruppen veranschaulicht die visuelle Darstellung, dass die Gruppenmitglieder in fast allen Runden Maximalbeiträge geleistet haben. Während in Gruppe 2 (NoP_G2) im Sinne von Fallucchi et al. (2019) alle vier Gruppenmitglieder als *unconditional high contributors* (UCH) bereits ab der ersten Runde immer 20 Punkte zum gemeinsamen Projekt beitragen (Ausnahme: Vpn⁸ 40 trägt in Runde 10 und in Runde 20 jeweils null Punkte bei), startet Gruppe 1 zunächst von niedrigerem Niveau (60 Punkte, bestehend aus Beiträgen von 20, 20, 12 und 8 Punkten). Anschließend werden auch in Gruppe 1 von allen Spielern in jeder Runde 20 Punkte beigegeben (bis auf Runde 20, in der eine Vpn einmalig null Punkte beiträgt). Obwohl diese Gruppe also mit drei Spielern des AF-Typs, nämlich zwei Spielern mit maximal altruistischem Frame von 20 Punkten und einem Spieler mit 12 Punkten, und einem Spieler mit egoistischem Frame (EF-Typ) 25 % unterhalb der maximalen Beitragshöhe gestartet ist, entscheiden sich ab Runde 2 alle vier Gruppenmitglieder zu maximaler Kooperation und sind damit als UCH im Sinne von Fallucchi et al. (2019) klassifizierbar. Gruppe 4 entspricht in der visuellen Analyse einer Gruppe *konditionaler Kooperativeure*, bei denen die individuellen Beiträge relativ schwach um den Gruppendurchschnitt streuen. Hier sind in der ersten Runde drei Vpn EF-Typen, und nur eine Vpn startet mit 20 Punkten in der ersten Runde (AF-Typ). In Gruppe 3 dagegen scheinen sich die Spieler nicht konsistent einer Strategie zuordnen zu lassen, wobei die Beiträge in dieser Gruppe nicht gegen null streben, sondern stark um den Mittelwert der gemeinsamen Beiträge streuen und zwischen null und 20

8 Die Nummerierung der Vpn orientiert sich am Codebook zum Datensatz von Betz, Esser, Biniossek (2020), um eine Vergleichbarkeit der Studien herzustellen.

Punkten hin und her springen, mit Ausnahme von Vpn 41, die konstant 20 Punkte beiträgt⁹. Diese Verteilung entspricht H1c.

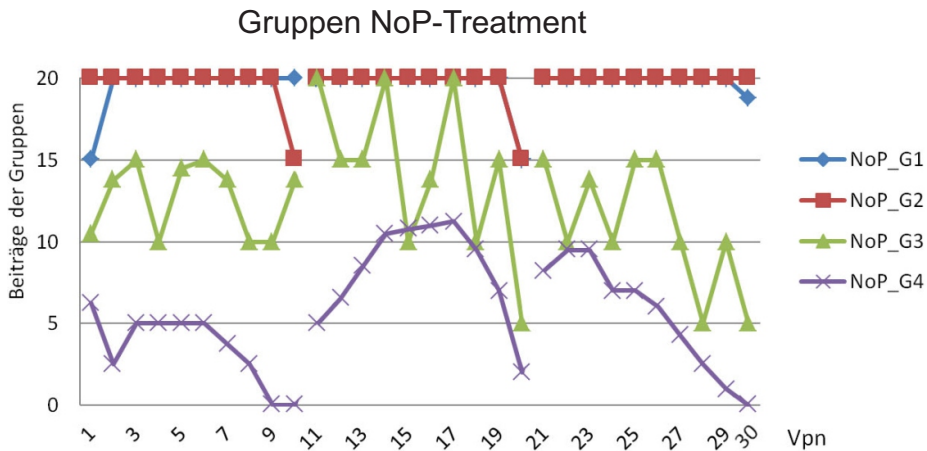


Abbildung 3 Durchschnittliche Gruppenbeiträge im NoP-Treatment

Nachfolgend wird überprüft, ob die visuell identifizierbaren Typen auch der Klassifizierung von FGF (2001) bzw. Thöni und Volk (2018) und Fallucchi et al. (2019) entsprechen. Die Beitragszahlungen in **NoP_G1** entsprechen den Kriterien für UCH (Fallucchi et al. 2019), allerdings kann in dieser Gruppe die Frage gestellt werden, ob Vpn 34 und Vpn 36 aufgrund ihres jeweiligen Erstrundenverhaltens eher als *strong conditional cooperators* (Fallucchi et al., 2019) eingeordnet werden müssen. Aufgrund der starken Signale in der ersten Runde gelingt hier den Spielern ab Runde 2 eine Koordination auf Maximalniveau. Die Beitragszahlungen in **NoP_G2** entsprechen über alle Runden den Kriterien der UCH. Die gesamten Beiträge von **NoP_G3** korrelieren hochsignifikant ($p = .002$) mit $|\rho| = .296$. Gemäß den Kriterien von Thöni und Volk (2018) qualifiziert diese Korrelation Gruppe 3 jedoch nicht als *konditionale Kooperateure*, da dann $|\rho| \geq .500$ sein müsste. Vpn 41 aus NoP_G3 trägt in allen Runden immer den vollen Betrag bei, ist also entsprechend der Definition von Fallucchi et al. (2019) ein UCH. Vpn 42 aus NoP_G3 trägt in 22 von 30 Fällen 20 Punkte zum gemeinsamen Projekt bei, einmal 18, einmal 15 und sechsmal null. Ob diese Vpn als *unkonditionaler Kooperator* im Sinne von Thöni und Volk (2018) klassifiziert werden kann, ist davon abhängig, wie streng die Definition bezüglich des konstanten

9 Kritisch ist hier anzumerken, dass trotz eines ähnlichen durchschnittlichen Beitragsniveaus die Anteile der unterschiedlichen Spielertypen im NoP sehr stark von den Anteilen der Spielertypen bei FGF (2001) abweichen. Während FGF (2001) von keinem einzigen Fall *unkonditionaler Kooperateure* bzw. UCH berichten, erfüllten im NoP-Treatment aufgrund der extrem positiven Beitragsentwicklung in zwei Gruppen insgesamt 50 % aller Spieler dieses Kriterium. Gruppenspezifische Verläufe scheinen zudem auch die starke Variation von Typen in der Literatur zu erklären (FGF, 2001; Thöni & Volk, 2018; Fallucchi et al., 2019).

Beitrags interpretiert wird. In dieser Gruppe kommt es aufgrund von wiederholten Störreizen (z. B. Beiträge gleich null) nie zu einer Koordination auf ein bestimmtes Gruppenverhalten. Die Beiträge der NoP_G4 korrelieren hochsignifikant mit dem Mittel der Beiträge der Vorrunde der anderen Spieler in dieser Gruppe. Vpn 45 lässt sich als *konditionaler Kooperator* mit $|\rho| = .634$, $p = .000$ klassifizieren. Dieser Spieler hatte in der ersten Runde null Punkte beigetragen. Vpn 46 liegt in ihrem Kooperationsniveau mit $|\rho| = .588$, $p = .001$ etwas unter Vpn 45, fällt aber noch in die Definition eines *konditionalen Kooperators* gemäß Thöni und Volk (2018). Vpn 47 verfehlt mit Pearson's Korrelationskoeffizient $|\rho| = .456$ und dem Signifikanzniveau $p = .013$ knapp das $|\rho| \geq 0.5$ -Level und ist damit nicht mehr als *konditionaler Kooperator* kategorisierbar. Vpn 48 erreicht mit $|\rho| = .689$, $p = .000$ den höchsten Kooperationswert in dieser Gruppe. H1c trifft damit zu.

Überprüfung von Hypothese 2

Reframing im NoP-Treatment

Um zu ermitteln, welche gedanklichen Modelle hinter den jeweiligen Typen stehen, werden nachfolgend analog zur Method of Difference (Lieberson, 1994; Lijphart, 1971; Mill, 1843) die Antworten zu den Entscheidungsgründen und zur Wahrnehmung einer Situationsänderung analysiert. Zunächst werden die Entscheidungsgründe für das Beitragsverhalten in der ersten Runde des NoP-Treatments betrachtet. Hierbei wird analog zu Esser (2018) zwischen Altruisten mit Altruismus-Frame (AF-Typ) und Egoisten mit Egoismus-Frame (EF-Typ) unterschieden, wobei die AF-Typen laut Definition in der ersten Runde 10 bis 20 Punkte geben und die EF-Typen weniger als 10 Punkte. Im NoP-Treatment sind in der ersten Runde sechs Vpn (37.5 %) als EF-Typen klassifizierbar. Wichtigste Entscheidungsgründe für Egoisten in der ersten Runde sind auf einer Skala von 1 bis 7 *Reziprozität* (6.3), *Entscheidungsunsicherheit* (5.5) und *Outcome-Maximierung* (4.7), während AF-Typen *Emotionalisierung* (5.0), *Fairness* (4.9) und *Outcome-Maximierung* (4.7) als wichtigsten Grund angeben. Damit waren AF-Typen fast im gleichen Ausmaß an *Outcome-Maximierung* orientiert wie EF-Typen. Diese unterscheiden sich von AF-Typen in erster Linie durch ihre höhere Reziprozitätsorientierung und ihre höhere Ambivalenz bezüglich der getroffenen Entscheidung (Item *Entscheidungsunsicherheit*).

Nachfolgend soll die Frage beantwortet werden, ob für die Vpn, die eine Veränderung der Situation im Spielverlauf wahrgenommen haben, auch ein geändertes Entscheidungsverhalten beobachtbar ist und ob sich die entsprechenden Entscheidungsgründe beispielsweise in Richtung der Gruppenkultur geändert haben. Dazu wurden die Vpn nach den drei Experimenten gefragt, ob sie eine Veränderung der Situation wahrgenommen haben, und anschließend die Verhaltensgründe *Outcome-Maximierung*, *Fairness*, *Reziprozität*, *Entscheidungsunsicherheit* und *Orientierung an den Mitspielenden* anhand einer 7-stufigen Skala erhoben. Für den Vergleich des Ver-

haltens nach der Wahrnehmung einer Situationsänderung und einer eventuellen Änderung der Entscheidungsgründe werden nachfolgend von den insgesamt 16 Vpn des NoP-Treatments lediglich die sechs Spieler betrachtet, die ihre Verhaltensstrategie gewechselt haben, und die Verhaltensentscheidungen den jeweils gemessenen, individuellen Entscheidungsgründen gegenübergestellt: die beiden Spieler aus der NoP_G1, die in der Erstrunde 8 bzw. 12 Punkte beigetragen haben (jedoch in den Folgerunden immer 20), sowie die Profile der vier Spieler aus der NoP_G4, die konditional kooperieren und in der ersten Runde 0, 0 oder 5 (EF-Typ) bzw. 20 (AF-Typ) Punkte beigetragen haben. Im Folgenden wird das Spielverhalten den jeweils gemessenen, individuellen Entscheidungsgründen gegenübergestellt. Die übrigen zehn Spieler werden nicht betrachtet, da bei den sechs UCH aus NoP_G1 und NoP_G2 sowie zwei Spielern aus NoP_G3 das Verhalten konstant bleibt und bei den weiteren zwei Spielern aus NoP_G3 weder konsistente Entscheidungsstrategien noch ein auf die Gruppenidentität zurückführbares Entscheidungsverhalten erkennbar ist.

Für UCH werden gemäß Fallucchi et al. (2019) starke, situationsüberdauernde prosoziale Motive, wenig Eigennutz und eine geringe *Orientierung an den Mitspielenden* prognostiziert. Die betreffenden Vpn aus NoP_G1 geben beide an, dass sich eine starke Änderung der Situation ab der zweiten Runde ergeben habe, jedoch ergibt sich daraus nur für den EF-Typ (8 Punkte Erstrundenbeitrag) eine starke Änderung der Gründe für sein Verhalten ab der zweiten Runde – der AF-Typ passt dagegen nur sein Verhalten ohne Änderung seiner Verhaltensgründe an. *Fairness* ist für den EF-Typ in der ersten Runde nicht wichtig, wird aber durch die geänderte Situation dominant (Skalenwerte 3/7)¹⁰, während der AF-Typ konstant stark an *Fairness* orientiert ist (7/7). Überraschend ist, dass für den EF-Typ die *Outcome-Maximierung* zunächst ebenfalls weniger wichtig ist, nach der wahrgenommenen Änderung der Situation aber wichtig wird (3/5), und der AF-Typ die *Outcome-Maximierung* konstant als sehr wichtigen Grund (7/7) sieht. Für den EF-Typ ist die *Orientierung an den Mitspielenden* extrem wichtig (6/7), für den AF-Typ nur in mittlerem Maße (4/4).

Der auffälligste Unterschied zu den vier *konditionalen Kooperatoren* aus NoP_G4 ist, dass dort bei drei der vier Spieler sowohl in der Erstrunde als auch nach der von drei der vier Spieler empfundenen Situationsänderung *Reziprozität* im Vordergrund steht (7/7; 7/7; 1/6; 7/4). *Fairness* scheint dagegen die Entscheidung weniger zu dominieren (6/5; 4/7; 7/4; 3/4). Überraschend ist zudem, dass die vier *konditionalen Kooperatoren* nicht stärker an *Outcome-Maximierung* orientiert sind (6/6; 7/7; 7/3; 1/4) als die beiden Spieler aus der NoP_G1. Jedoch weisen in NoP_G4 alle Spieler erwartungsgemäß eine mittlere bis starke *Orientierung an den Mitspielenden* (4/6; 7/7; 7/6; 4/4) auf.

Die oben dargestellten Ergebnisse bezüglich H1 und H2 im NoP-Treatment verdeutlichen die Gravitationskraft der gruppenspezifischen Entwicklung. In drei der vier Gruppen finden die Beiträge zu einem Gleichgewicht, das in den NoP_G1 und

10 In den Klammern stehen als erstes die Werte aus der Befragung für das Verhalten in der ersten Runde, dahinter stehen die Werte für die Entscheidungsgründe, nachdem subjektiv eine Veränderung der Situation wahrgenommen wurde.

NoP_G2 beim Maximalwert 20 und in NoP_G4 in der konditional kooperativen Gruppenkultur liegt. In NoP_G1 und NoP_G4 befinden sich in Runde 1 zunächst unterschiedliche Typen, die erst ab der zweiten Runde zum jeweiligen, gruppenspezifischen Typ konvergieren. Die Verläufe werden nicht von konditionaler Kooperation geprägt, die von *Trittbrettfahrern* Richtung null gezogen wird, sondern sind davon bestimmt, wie stark Individuen an ihrer eigenen Strategie festhalten oder sich an der Gruppe orientieren. Dies wird an NoP_G3 mit extremen Ausschlägen deutlich, in der zwei Spieler nicht von ihren antagonistischen Strategien abrücken und die beiden gruppenorientierten Spieler orientierungslos zurücklassen. Daher besteht hier weder Konvergenz der Beiträge noch Konvergenz der Typen – H2 trifft somit für alle Gruppen des NoP-Treatments zu, außer für NoP_G3.

Deskriptive Analyse und Reframing im NoF-Treatment

Das Treatment ohne Feedback (NoF) unterscheidet sich substantiell von den herkömmlichen Treatments im experimentellen ÖGS und somit auch vom NoP-Treatment in der vorliegenden Arbeit. Im NoF-Treatment werden in den Runden 2 bis 10 (**NoF-Exp1**) ($n = 144$) durchschnittlich 7.83 Punkte beigetragen ($\text{Median} = 5$, $SD = 8.37$). Dabei wird in 40.3 % der Fälle nichts zum gemeinsamen Projekt beigetragen, in 26.4 % wird ein Beitrag von 20 und in 10.4 % ein Beitrag von 5 Punkten gewählt. In den Runden 11 bis 20 (**NoF-Exp2**) ($n = 160$) werden mit durchschnittlich 6.9 ($\text{Median} = 1$) niedrigere, jedoch ebenfalls stark variierende ($SD = 8.5$) Beiträge geleistet. In 47.5 % der Fälle wird kein Beitrag geleistet, in 22.5 % ein Beitrag von 20 und in 6.9 % ein Beitrag von 5 Punkten. Nachdem die Spieler nach Abschluss des zweiten Experimentes zum ersten Mal ein Feedback über die Beiträge der anderen Mitglieder ihrer Gruppe erhalten, sinken die durchschnittlichen Beiträge für die Runden 21 bis 30 (**NoF-Exp3**) auf durchschnittlich 3.14 Punkte ($\text{Median} = \text{Null}$, $SD = 5.71$), jedoch ergeben sich keine Änderungen der drei am häufigsten geleisteten Beitragshöhen (61.9 % null, 10 % fünf und 6.9 % 20 Punkte).

Nach dem Verhalten in der ersten Runde sind in der Klassifikation Essers (2018) jeweils die Hälfte der Spieler dem AF-Typ (davon sechs Spieler mit einem Beitrag von 20 und zwei Spieler mit einem Beitrag von 15 Punkten) und dem EF-Typ (dreimal null, einmal 3, dreimal 5 und einmal 7 Punkte Beitrag) zuzuordnen. Bei Klassifikation der Verhaltenstypen fällt zunächst auf, dass nur ein Spieler entsprechend standardökonomischer Prognose bzw. einer engen Definition der RCT konstant null Punkte beiträgt (jedoch trägt selbst dieser Spieler in Runde 29 des zweiten Experimentes 9 Punkte zum Projekt bei). Während zwei weitere Spieler ebenfalls durchschnittlich weniger als einen Punkt beitrugen und daher als *Trittbrettfahrer* klassifiziert werden können, leisten zwei Spieler konstant über alle 20 Runden einen Beitrag von 20 Punkten und sind als UCH klassifizierbar. Eine weitere Vpn trägt in den Runden 1 bis 9 sowie 10 bis 19 immer 20 Punkte bei und in den Runden 10 und 20 jeweils null Punkte – dies wäre als Endrundeneffekt klassifizierbar, wenn ein Feedback gegeben worden wäre. Dennoch ist auch diese Vpn als UCH klassifizierbar.

Die zehn verbleibenden Vpn entsprechen keinem der in der Literatur genannten Verhaltenstypen. In dieses Bild passt, dass bei rein visueller Betrachtung auf der Aggregatebene (Abbildung 2) die Beiträge wie im typischen ÖGS mit Feedback im Rundenverlauf sinken. In der Literatur werden sinkende Verläufe im ÖGS als Beleg für konditionale Kooperation interpretiert. Die Erklärung sinkender Beiträge im ÖGS als Interaktionseffekt zwischen *konditionalen Kooperatoren* und *Trittbrettfahrern* erscheint jedoch nicht hinreichend, da die Beiträge im NoF-Treatment auch ohne Feedback in ähnlichem Ausmaß sinken. Zudem besteht bei einem Mittelwertvergleich dieser beiden Stichproben kein Unterschied. Lediglich der inverse Schlussrundeneffekt in NoF Runde 20 stört diesen Gesamteindruck.

Reframing

Abgesehen von drei UCH, die konstant 20 Punkte beitragen und drei *Trittbrettfahrern*, die (fast) immer null Punkte geben, ist das Beitragsverhalten der übrigen zehn Spieler erratisch und scheint keinen erkennbaren Mustern zu folgen, die bei Annahme einer zentralen Exekutive und rationalen Verhaltens zu erwarten wären. Mehrheitlich existieren in den ersten 20 Runden keine stabilen Verhaltenstypen, die ohne Feedback zu erwarten gewesen wären. Der Einbruch der Beiträge in Runde 21 nach erfolgtem Feedback entspricht den Vorhersagen konditionaler Kooperation, nicht jedoch das anschließend erneut erratische Verhalten von sieben der 16 Spieler, aus dem in den Runden 22 bis 30 ein kollektiv steigender Beitragsverlauf resultiert.

Elf der 16 Vpn gaben an, eine Änderung der Spielsituation wahrgenommen zu haben, davon eine Vpn nach der ersten Runde, als das Feedback über die eigenen Einkünfte und das Feedback über die Beiträge der anderen Spieler ausgeblieben war, eine Vpn nach der ersten und der 20. Runde, eine Vpn nach der 10. Runde, eine Vpn nach der zehnten und der 20. Runde (nach der erstmals ein Feedback gegeben wurde), drei Vpn nach der 20. Runde, zwei Vpn in der 30. Runde sowie zwei Vpn zu einem nicht näher spezifizierten Zeitpunkt. Zwei der drei UCH hatten, ausgelöst durch das Feedback, eine Veränderung der Spielsituation wahrgenommen.

Bei Vpn 66 führte das Feedback zu einer anschließend viel größeren Bedeutung von *Reziprozität* (4/7) sowie einer geringeren *Entscheidungsunsicherheit* (7/1). Bei Vpn 75 ergaben sich demgegenüber große Veränderungen bei *Fairness* (6/1) als angegebene Verhaltensgrund (*Reziprozität* bleibt bei 1 gleich). Vpn 72, die konstant in allen 30 Runden immer 20 Punkte zum gemeinsamen Projekt beiträgt, hat laut eigener Aussage die viel geringeren Beiträge ihrer Mitspielenden registriert. Jedoch hat sich für sie persönlich durch das Feedback keine Veränderung ergeben und sie hat folgerichtig für ihr Verhalten nach der objektiven Änderung dieselben Verhaltensgründe genannt. Vpn 72 begründet dies damit, dass ihr von Anfang an bewusst gewesen sei, dass sie ausgenutzt werde, sie jedoch auch nach dem Feedback ihrer Grundhaltung habe treu bleiben wollen: „Mir war von Anfang an klar, dass es Leute geben wird, die nicht alles investieren, jedoch wie schon beschrieben war es mir wichtiger mich ‘richtig’ zu Verhalten, als mit mehr Geld nach Hause zu gehen.“ Bei den vier weiteren Vpn, die überhaupt eine Veränderung der Situation durch das Feedback zwischen Runde 20

und Runde 21 wahrgenommen haben, ergeben sich durchschnittlich nur geringfügige Änderungen der Verhaltensgründe. Zusammengefasst bestehen im NoF-Treatment nur relativ geringe Zusammenhänge zwischen Persönlichkeitstyp, Verhaltenstyp und geäußerten Entscheidungsgründen. H2 kann auch für das NoF-Treatment bekräftigt werden, da eine Mehrheit der Vpn teils mehrmals ihr Verhaltensmuster und ihren Typ wechselt.

Hypothese 3

Nach der RCT, aber auch nach dem Konzept der *home-grown beliefs* von FGF (2001), sollten die individuellen Beiträge im NoF-Treatment im Rundenverlauf nicht variieren, sondern sich ab der zweiten bis zur 20. Runde nicht mehr grundsätzlich verändern – vor allem aber nicht steigen. Im dritten Experiment (NoF-Exp3) sollten sich nach dem Feedback über die Beiträge der anderen Spieler die eigenen Beiträge entweder im Sinne konditionaler Kooperation an den Gruppendurchschnitt angleichen oder aus dem Eigennutzmotiv bzw. aus Enttäuschung über die im Vergleich zu den eigenen Beiträgen zu niedrigen Beiträge der anderen Spieler in den Vorrunden nun in Richtung null bewegen und anschließend in den Runden 22 bis 30 konstant bleiben, jedoch ebenfalls auf keinen Fall mehr steigen. So erheben z. B. FGF (2001), Thöni und Volk (2018) und Fallucchi et al. (2019) insoweit den Anspruch, mit ihren Typologien das Verhalten der meisten Spieler erklären zu können. Für keine Variante der RCT ist die Prognose ableitbar, dass die Beiträge einer Vpn im Treatment ohne Feedback in einer Folgerunde steigen. Insbesondere dürften Beiträge ab Runde 3 weder über dem Niveau der ersten Runde liegen, die von *home-grown beliefs* geprägt wird, noch über den Beiträgen der zweiten Runde, in der spätestens deutlich ist, dass kein Feedback erfolgt. Diese Prognosen sind nur aus der RCT, nicht aber aus Rahmentheorien ableitbar. Deshalb wird im Folgenden für das NoF-Treatment die Leistungsfähigkeit des RCT-Paradigmas sowie unterschiedlicher Typisierungen untersucht. Hierzu werden auf Individualebene die Verhaltensmuster, die Übereinstimmung von Verhalten zu den genannten Gründen sowie die Theorieferne bzw. Theorienähe der angegebenen Erklärungen betrachtet und die Schwere des Verstoßes gegen die Prognosen der RCT aufgezeigt.

Drei Vpn des Typs unconditional high contributor

Drei Spieler im NoF-Exp1 leisten in jeder Runde maximale Beiträge von 20 Punkten und lassen sich dem Verhaltenstyp UCH zuordnen. In NoF-Exp2 erweist sich diese Typisierung wie von Fallucchi et al. (2019) prognostiziert als höchst leistungsfähig, da alle Vpn, die in Experiment 1 als UCH klassifiziert wurden, weiterhin konstant hohe Beiträge von 20 Punkten leisten. In NoF-Exp3 versagt diese Theorieferne, nur am Ziel der trennscharfen Identifizierung ähnlicher Entscheidungsmuster orientierte Typisierung, da zwei Akteure zu *Trittbrettfahrern* werden und nur noch ein Spieler als UCH klassifizierbar ist. Dabei handelt es sich beim Verhalten dieses UCH, wie oben dargestellt, um eine selbstorientierte ethische Grundhaltung. Demgegen-

über kann die Klassifikation von FGF (2001) bereits ab der zweiten Runde von NoF-Exp1 nicht bestätigt werden, da in dieser Typologie UCH nicht vorgesehen sind und ab der zweiten Runde der Endrundeneffekt prognostiziert wird, weil kein Spieler für Trittbrettfahrerverhalten negative Konsequenzen befürchten muss und dieses Verhalten daher auch von seinen Mitspielern erwarten sollte. Das Verhalten aller drei Individuen weist eine relative Nähe zum Konstrukt UCH auf und ist weder durch die RCT noch durch konditionale Kooperation erklärbar.

Trittbrettfahrer versus own maximiser

Vpn_78 trägt in der ersten Runde 15 Punkte und in der zweiten Runde 10 Punkte bei und leistet von Runde 3 bis 30 keine Beiträge mehr. Da keine klaren Vorgaben bestehen, wie Vpn_78 theoretisch fundiert und empirisch begründet klassifizierbar ist, werden nachfolgend formale und inhaltliche Argumente gegenübergestellt. Technisch entspricht Vpn_78 dem Verhaltenstyp *own maximiser* (Fallucchi et al., 2019). Die Definition von *Trittbrettfahrer* (FGF, 2001) erlaubt anders als *own maximiser* nur Beiträge in Höhe von null, während *own maximiser* als Kriterium lediglich eine modale Allokation von null verlangt. Damit wäre die formal als *own maximiser* klassifizierte Vpn_78 streng genommen nicht kompatibel mit der RCT. Wird jedoch inhaltlich argumentiert, könnte Vpn_78 möglicherweise als Verhaltenstyp *konditionaler Kooperator* interpretiert werden, der zunächst kooperiert und anschließend bei Ausbleiben des Feedbacks Beiträge entsprechend seiner Erwartung über die Höhe der Beiträge seiner Mitspielenden leistet. Wird diese inhaltliche Argumentation gewählt, sind aus dem Verhaltenstyp *konditionaler Kooperator* für das NoF-Treatment zwei Prognosen plausibel ableitbar: Konstant hohe Beiträge ab Runde 1 entsprechend den *home-grown beliefs*, da kein Störreiz auftritt, sowie konstante Beiträge ab Runde 2, die jedoch oberhalb von null liegen müssen, da konditional davon ausgegangen werden muss, dass andere Spieler durchschnittlich Beiträge leisten, die größer als null sind. Beide Argumentationen greifen bei Vpn_78 jedoch nicht, da (mit Verzögerung) ab Runde 3 keine Beiträge geleistet werden. Vpn_74, die in 29 Runden keine Beiträge leistet und lediglich in Runde 14 schwach gegen die Prognosen der RCT verstößt, kann ebenfalls als *own maximiser* (Fallucchi et al., 2019) eingestuft werden. Wird in geringem Ausmaß ein Verstoß gegen grundlegende Annahmen der RCT toleriert, ließe sich diese Vpn auch als *Trittbrettfahrer* mit der RCT kompatibel typisieren.

11 Vpn, die keinem Typ zuordenbar sind

Von den 16 Spielern im NoF-Treatment lassen sich drei Spieler als UCH klassifizieren und werden aus der nachfolgenden Analyse ausgeschlossen. Vpn_78 trägt in der ersten Runde 15, in der zweiten Runde 10 und danach konstant null Punkte zum gemeinsamen Projekt bei. Sie lässt sich damit zwar keinem Typ zuordnen, jedoch ist ihre Verhaltensstrategie mit einer weiten Version der RCT vereinbar, sodass sie ebenfalls nicht berücksichtigt wird. Von den beiden *own maximisern* wird Vpn_74 ebenfalls nicht berücksichtigt, da sie in einer Runde nur schwach gegen die RCT

verstößt, jedoch wird Vpn_76 einbezogen, da sie in einer Runde fast die Hälfte ihrer Ausstattung beiträgt. In Abbildung 4 ist dargestellt, in wie vielen Runden von den in die Analyse einbezogenen elf Vpn grundlegend gegen die RCT verstoßen wird.

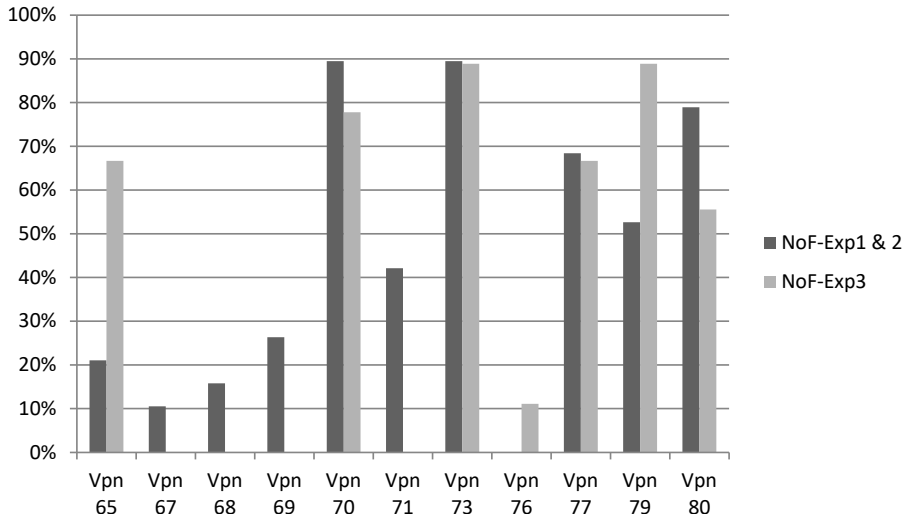


Abbildung 4 Prozentualer Anteil der Runden mit höheren Beiträgen als in den jeweiligen Vorrunden (Vergleich [NoF-Exp1 & NoF-Exp2] mit [NoF-Exp3])

In der Spielvariante ohne Feedback lässt sich das Verhalten von zehn der elf Vpn nicht durch die RCT-Kriterien erklären, wobei dies bei fünf dieser zehn Vpn in mindestens der Hälfte der betrachteten Runden 2 bis 20 nicht möglich ist. Nach dem Feedback kommt es bei isolierter Betrachtung von **NoF-Exp3** (Runden 22 bis 30) bei insgesamt sieben der elf Vpn entgegen den Prognosen der RCT zu höheren Beiträgen als in Vorrunden, bei sechs dieser sieben Vpn in über der Hälfte der Runden. Insgesamt entspricht das Verhalten von keiner der elf Vpn den Grundannahmen der RCT oder den Typologien von FGF (2001), Thöni und Volk (2018) und Fallucchi et al. (2019). In 135 (43.8 %) der 308 untersuchten Entscheidungen haben die elf Vpn grundlegend gegen die RCT verstoßen, davon in 94 (45 %) Runden von NoF-Exp1 und NoF-Exp2 sowie in 41 (41.4 %) Runden von NoF-Exp3. Nach Schwere des Verstoßes kategorisiert¹¹ wurde insgesamt ([NoF-Exp1 & NoF-Exp2]/[NoF-Exp3]) 19 (13/6) Mal das Kriterium von B3, 25 (14/11) Mal B2, 23 (13/10) Mal B1 sowie 22 (16/6) Mal A3, 5 (4/1) Mal A2 und 41 (34/7) Mal A1 verletzt. Auffällig ist die Heterogenität, wenn

11 (A1) in mindestens einer der vorausgehenden Runden niedrigere Beiträge; (A2) Beiträge, die über den Beiträgen der unmittelbar vorangehenden Runde liegen; (A3) Fallen der Beiträge von Vor-Vorrunde auf Vorrunde mit anschließendem Steigen. Einen Sonderfall stellen Beiträge (B1) oberhalb der Beiträge der ersten Runde dar. Ebenfalls der Guttman-Skala-Logik folgend: (B2) Beiträge oberhalb der Beiträge der unmittelbar vorausgehenden Runde; (B3) Fallen der Beiträge von Vor-Vorrunde auf Vorrunde mit anschließendem Steigen.

es darum geht, in welchem Experiment in welchem Ausmaß das Verhalten der Vpn nicht durch die RCT-Kriterien erklärbar ist. Der erneute Anstieg von Beiträgen nach einem unmittelbar vorangegangenen Sinken (Kriterium A3 und B3) ist bei neun Vpn zu beobachten, bei fünf dieser Vpn sowohl in den ersten 20 Runden als auch in den letzten 10 Runden. Beiträge, die über dem Betrag der ersten Runde liegen, können bei zehn Vpn beobachtet werden. In NoF-Exp1 verletzen sechs Vpn dieses Kriterium. Ausschließlich bei diesen Spielern kommt es in NoF-Exp2 (4 Vpn) erneut zu Verstößen gegen dieses Kriterium. In NoF-Exp3 leisten sieben Vpn höhere Beiträge als in Runde 21. Zusammenfassend bekräftigt dies **H3**, da sich das Verhalten dieser elf betrachteten Vpn nicht anhand der RCT prognostizieren lässt. Wenn Prognosen ausschließlich auf die RCT fixiert sind, ist diese Vorgehensweise dafür verantwortlich, dass das Verhalten der oben analysierten Vpn nicht systematisch erfasst und erklärt werden kann.

Diskussion

Zwischen dem Experiment von F&G (2000), dem NoP-Treatment und dem NoF-Treatment bestehen in der ersten Runde keine signifikanten Mittelwertunterschiede (Bestätigung **H1a**), jedoch liegt das Beitragsniveau wie nachfolgend diskutiert sowohl von NoP als auch von NoF oberhalb der von F&G (2000) gemessenen Werte. Im Vergleich zum Erstrundenverhalten befinden sich in NoP und NoF mehr AF-Typen als EF-Typen (**H1b** wird nicht bestätigt). Die aggregierten Entscheidungsmuster zwischen F&G (2000) und dem NoP-Treatment sind strukturell ähnlich (kooperativer Beginn, kurzzeitiges Steigen in der zweiten Runde, dann im Rundenverlauf fallende Beiträge sowie ein Schlussrundeneffekt), während im NoF-Treatment nach ebenfalls kooperativem Start teils inverse kollektive Muster beobachtbar sind (unmittelbares Sinken auf ein niedrigeres Niveau, inverser Schlussrundeneffekt). Im NoP-Treatment kommt es in allen vier Gruppen zu gruppenspezifischen Beitragsverläufen und Anteilen der Typen (Bestätigung **H1c**), in drei der vier Gruppen bildet sich hierbei zudem eine gruppenspezifische Kultur mit gruppenspezifischen mentalen Modellen, Entscheidungsgründen und Verhaltenstypen. Bei den Spielern, die sich an diese gruppenspezifische Kultur anpassen, ist ein entsprechender Wechsel (*Reframing*) von Typen, Episoden und mentalen Schemata identifizierbar (Bestätigung **H2**). Demgegenüber bleiben im NoF-Treatment aufgrund des ausbleibenden Feedbacks die meisten Spieler orientierungslos, die Verhaltensgründe korrespondieren nur relativ schwach mit dem Verhalten und nur sechs der 16 Vpn sind hier einem Typ zuordenbar. Die individuellen Beiträge der meisten Spieler sind nicht konstant, sondern schwanken stark. Das Verhalten von elf Vpn lässt sich durch die RCT nicht erklären. Unter Ausschluss der ersten und der 21. Runde kommt es bei diesen Spielern in den 308 betrachteten Beitragsentscheidungen in 135 (43.8 %) Runden zu individuellen Beiträgen, die der RCT grundlegend widersprechen, jedoch im Einklang mit Rahmenmodellen wie dem MFS oder dem BRAC stehen, die beispielsweise bei ausbleibendem Feedback

von starken mentalen Prozessen und der möglichen Instabilität situativ aktivierter Episoden und Verhaltenstypen ausgehen (Bestätigung H3).

Im NoF- und im NoP-Treatment werden nicht signifikant höhere Erstrundenbeiträge geleistet als bei F&G (2000). Die kollektiv höheren Beiträge sind eine Folge des signifikant kooperativeren Verhaltens von Männern in der ersten Runde, wobei im NoF-Treatment dieser Unterschied im Beitragsverhalten zwischen Männern und Frauen auch in den Folgerunden in etwa konstant bleibt. Auch wenn Andreoni und Vesterlund (2001) anhand einer Metaanalyse für das Diktatorspiel bei Männern höhere Beiträge und ein starkes Festhalten an Strategien nachweisen, wenn es um geringe Beträge geht, jedoch niedrigere Beiträge bei hohen Einsätzen, besteht dennoch in der Literatur (z. B. Ledyard, 1995; Zelmer, 2003) weitgehender Konsens, dass im ÖGS kein systematischer Gendereffekt zu erwarten ist.

Nachfolgend werden die methodischen und theoretischen Untersuchungsergebnisse zueinander in Beziehung gesetzt, um zu erklären, warum die oben diskutierten, zunächst nicht signifikant höheren Erstrundenbeiträge kaum Einfluss auf die Ergebnisse im NoF-Treatment haben, jedoch im NoP-Treatment über eine Verkettung methodischer und psychologischer Faktoren dazu beitragen, dass sich die Beiträge im Verlauf der Runden auf Aggregatebene immer weiter von den Verläufen bei F&G (2000) entfernen (Abbildung 2). Nach sinkenden Beiträgen in der zweiten und in der dritten Runde schwankt das Kooperationsniveau im NoF-Treatment um ein Plateau. Die aggregierten Beiträge ergeben sich aus zwei unterschiedlichen Arten, wie die Spieler auf die (unklare) Situation im NoF-Treatment reagieren: Bei einem Drittel der Vpn besteht ein starker *Match* bezüglich der Wahrnehmung der Situation, der abgerufenen Episode und der selektierten Verhaltensstrategie (konstant null, konstant Maximalbeiträge oder nach zunächst kooperativem Start aufgrund ausbleibenden Feedbacks null), die konsistent beibehalten werden. Bei zwei Dritteln der Vpn besteht nur ein schwacher *Match* bezüglich der Interpretation der Spielsituation und der Angemessenheit der abgerufenen Episode. Hier kommt es zu unklaren, teils mehrmals sinkenden und erneut ansteigenden Beiträgen (die sich als *Match* individueller Heuristiken und mehrmaliges *Reframing* von Strategien interpretieren lassen) sowie beispielsweise zum Abruf konditionaler Kooperation als Episode, obwohl dies objektiv nicht zur Situation passt. Hieraus lässt sich als Kernaussage ableiten, dass nur ein starker *Match* zu Gleichgewichten führt (Bereich der hohen Leistungsfähigkeit der RCT). Ein schwacher *Match* führt jedoch zu starken inneren Prozessen, einer Orientierung an schwachen (subjektiv salienten) Signalen oder (falls möglich) an den anderen Spielern und zum Abruf heterogener Episoden und Muster (Bereich der hohen Leistungsfähigkeit von Rahmentheorien).

Im **NoP-Treatment** führt demgegenüber der gruppenspezifisch eingeschlagene Pfad zu den von Rahmentheorien prognostizierten Anpassungsprozessen, gruppenspezifischen Gleichgewichten und der Harmonisierung von Typen. Die Bildung einer gruppenspezifischen Kultur in drei der vier NoP-Gruppen trägt dazu bei, dass sich der Abstand der im Vergleich zu F&G (2000) zunächst nur moderat höheren Erstrundenbeiträge in der zehnten Runde versiebenfacht. Der strukturell bedingte Timelag ist wesentlich für die üblicherweise sinkenden Beiträge im ÖGS verant-

wortlich – alle drei Gruppen, in denen sich eine gruppenspezifische Kultur gebildet hat, setzen dem durch die Spielstruktur bedingten Absinken der Beiträge nachhaltig über alle drei Experimente ihr gruppenspezifisches Gleichgewicht entgegen und können damit sinkende Beiträge vermeiden.

Auf Individualebene entfalten die RCT und die Typisierungen nur eine geringe Erklärungskraft, wenn die Definition einer Situation unklar ist und Individuen lediglich einen schwachen *Match* aufweisen. Hier werden Episoden abgerufen, die in dieser unsicheren Situation verfügbar und passend sind: Es kommt häufig zu instabilen Verhaltensmustern, Handlungsskripten, einem mehrfachen *Reframing* im Zeitverlauf und großer Heterogenität zwischen den Individuen. Aus der Perspektive von Rahmentheorien stellt die RCT einen Spezialfall dar, der aus vielen Gründen dennoch eine besonders hohe Erklärungskraft besitzt. Aus lerntheoretischer Perspektive wirkt eine Orientierung am Nutzenmaximierungsprinzip kognitiv entlastend und führt oftmals zu höheren Verstärkerwerten. Die Grundprinzipien der RCT werden zudem über Sozialisationsprozesse erlernt und konsistentes, nutzenmaximierendes Verhalten ist oftmals kulturell erwünscht. Damit ist rationales Handeln für viele Individuen in unklaren Situationen ein starkes gedankliches Modell. Starke Signale, Situationen, Leitideen oder Individuen mit einem starken *Match* entfalten aus dynamischer Perspektive eine große Anziehungskraft auf Individuen mit schwachem *Match*, die nach Orientierung suchen. Die Anziehungskraft der RCT-orientierten Spieler auf andere, unentschlossene Spieler ist somit ein wichtiger Faktor für die große Erklärungskraft der Theorie auf Aggregatebene und trägt zu Gleichgewichten bei, die auf der Makroebene stabil sind. Zudem besteht sowohl bei Vpn als auch bei Experimentatoren die Tendenz, gezeigtes Verhalten zu rationalisieren, jedoch besteht bei der Gesamtschau aller Verhaltensmuster und Entscheidungsgründe von Individuen eine erhebliche Ferne zu Prinzipien der RCT¹².

Limitationen der Arbeit sind aus methodischer Perspektive die geringe Fallzahl und die nicht-repräsentative Stichprobe mit einem Gender-Bias. Da das Experiment eine Replikation¹³ des F&G (2000) Experimentes darstellt, ist das relativ höhere Kooperationsniveau wahrscheinlich nicht auf das Spieldesign, sondern auf die Stichprobe zurückzuführen. Als weitere, grundsätzliche Limitationen werden in der Literatur die externe Validität von Laboruntersuchungen, die Übertragbarkeit auf reale öffentliche Güter sowie die Korrespondenz von gemessenen Einstellungen und Verhalten diskutiert. Für die vorliegende Arbeit sind diese Einwände jedoch von eher geringer Relevanz, da sowohl die RCT mit den aus ihr ableitbaren Typi-

12 In beiden Treatments werden Gründe genannt, die dem gezeigten Entscheidungsverhalten widersprechen (*Fairness* als wichtiger Verhaltensgrund bei Beiträgen von null). Einige Spieler berichten von einem Wechsel ihrer Motive, ohne dass eine entsprechende Änderung des Verhaltens beobachtbar ist. Andere Spieler ändern demgegenüber ihre Verhaltensstrategien, ohne eine geänderte Situation wahrzunehmen.

13 Die exakte Replikation umfasst sowohl den Inhalt und die Form der Instruktionen als auch die Verwendung der Original-Programmierung für die ersten 10 Spielrunden des NoP-Treatments.

sierungen als auch Rahmentheorien den Anspruch einer General Theory of Action erheben, die für jedes Individuum in jeder Situation an jedem Ort gelten.

Als Ausblick sollen nachfolgend methodologische Implikationen für spieltheoretische Experimente skizziert werden. Die experimentelle Spieltheorie geht davon aus, dass mit ausreichend hohen Auszahlungen für das normative ÖGS auch die entsprechenden Nutzenfunktionen sowie der Charakter als Kooperationsspiel induziert werden. Individuen entscheiden anschließend entsprechend ihrer (sozialen) Präferenzen. Die vorliegende Untersuchung zeigt, dass die Spielsituation empirisch sehr heterogen wahrgenommen wird und sich sowohl Präferenzen als auch Typen als abhängige Variablen im Rundenverlauf ändern. Das ÖGS kann bezüglich seiner Identität unterschiedliche Wahrheitswerte annehmen (Kooperationssituation, Wettbewerbssituation, Gruppeneffizienz), so wie dies simultan je nach Kontext auch für Mitspieler gelten kann, die willkürlich als Freunde, Kommilitonen oder Konkurrenten kategorisiert werden können. Wird nun angenommen, dass ein Attribut mehr als zwei Wahrheitswerte annehmen kann, ließe sich eine unendlichwertige Spieltheorie (z. B. aufbauend auf Łukasiewicz-Logiken L_v) formulieren, um das empirische ÖGS abzubilden.

Fazit

In der vorliegenden Untersuchung konnten auf Aggregatebene sowohl die anfänglichen Beiträge in der ersten Runde als auch die Struktur im Rundenverlauf des Experimentes ohne Bestrafungsoption von Fehr und Gächter (F&G) (2000) repliziert werden (NoP-Treatment). In Übereinstimmung mit der aktuellen Literatur wurde jedoch entgegen F&G (2000) auch der Typ *unconditional high contributor* (Fallucchi et al., 2019) nachgewiesen. Die Unterschiede zu F&G (2000) ergeben sich im Wesentlichen aus dem prägenden Einfluss der gruppenspezifischen Verläufe des Spiels. In den meisten Gruppen setzte sich eine gruppenspezifische Kultur durch, die bei Individuen mit schwachem *Match* zu einem Wechsel ihrer Verhaltensstrategie und damit des Typs führt. Hieraus leitet sich ab, dass jede Gruppe ein eigenes Treatment darstellt und eine Aggregation unterschiedlicher Gruppen methodologisch unzulässig ist. Bei Wiederholung des Experimentes konnte der in der Literatur berichtete Restart-Effekt auch für das NoP-Treatment nachgewiesen werden. F&G (2000) erklären die steigenden Beiträge bei Wiederholung ihres Experimentes damit, dass sie eine Bestrafungsoption eingeführt haben. Wenn jedoch auch ohne Bestrafungsoption die Beiträge steigen (isolierter Restart-Effekt), müssen neben Bestrafung weitere Mechanismen (z. B. das mentale Modell, der Gruppe nochmals eine Chance zu geben) für den Gesamteffekt verantwortlich sein. Da im NoF-Treatment kein Feedback gegeben wird, entfällt die koordinierende Wirkung der Gruppen, dennoch bleibt das Kooperationsniveau überraschend hoch, obwohl bereits ab der zweiten Runde der Endrundeneffekt wirken sollte, da keinerlei Konsequenzen für Trittbrettfahrerverhalten existiert. Aufgrund der uneindeutigen Situation zeigt die Mehrheit der Vpn

erratisches Verhalten, das weder durch Typen noch durch die RCT erklärbar ist, zudem wechseln auch hier die Verhaltensmuster, wodurch die Existenz situativer Typen und Persönlichkeitszüge bestätigt wird. Unsere Experimente zeigen, dass eine Vielzahl von Typen vorliegen kann, die Individuen aktiv auswählen und die sich im Zeitverlauf dynamisch ändern können. Rahmenmodelle stellen ein leistungsfähiges Gerüst dar, in das insbesondere bei unklaren äußeren Signalen alle Befunde eingepasst werden können. Das bedeutet zusammenfassend: Das individuelle Verhalten wird nicht von einer zentralen Exekutive bestimmt, die auf eine gegebene Spielsituation und stabile Persönlichkeits- und Präferenzstruktur nutzenmaximierend reagiert, sondern von der subjektiven Perzeption der objektiv vorgegebenen Struktur (Spielsituation, Spieldesign) und der Stärke des *Matches* des individuellen Zielsystems (Gruppenorientierung, Informationssuche, Gerechtigkeitsprinzipien) mit den situativen Stimuli. Als Fazit konnte das Potenzial und die große Prognosekraft von Rahmentheorien aufgezeigt werden, die die RCT in sich integrieren können. Der Verhaltenstyp *konditionaler Kooperator* ist über seine reziproke Entscheidungsstrategie definiert. Die technische Definition *konditionaler Kooperator* steht nur in einer indirekten Verbindung zu der RCT. Aus dieser Definition lassen sich keine Aussagen über die Höhe der Kooperation oder soziale Präferenzen ableiten. Konditionale Kooperation ist als definiertes Verhaltensmuster sowohl bei verzögertem Feedback im NoP-Treatment als auch ohne Feedback im NoF-Treatment identifizierbar. Damit erweist sich das hypothetische Konstrukt einerseits als leistungsfähig, andererseits ist das Verhaltensmuster *konditionaler Kooperator* auch in Situationen beobachtbar, in denen konditionale Kooperation objektiv nicht möglich ist, und stellt in diesem Gegenstandsbereich ein Artefakt mit einem beschränkten Geltungsbereich dar. Wird konditionale Kondition jedoch inhaltlich als Reziprozitätsmotiv interpretiert, dann stellt sie einen wichtigen Grund für das Entscheidungsverhalten dar, da dieses Motiv empirisch in gewissem Maße ausgeprägt ist. Einschränkend zeigt sich jedoch empirisch kein starker Zusammenhang zwischen dem gemessenen Reziprozitätsmotiv und dem identifizierten Verhaltenstyp *konditionaler Kooperator*. Es wird daher empfohlen, die gesamte Bandbreite der vielfältigen Untersuchungsstränge des Mixed-Methods-Designs (MMD) für die Typenbildung einzubeziehen.

Aus den vorliegenden Untersuchungsergebnissen ergibt sich eine Vielzahl neuer Forschungsfragen. Die zukünftige Forschung sollte insbesondere das individuelle Entscheidungsverhalten und die Entscheidungsgründe auf der Individualebene betrachten (kleine Granularität) und Instrumente entwickeln und anwenden, um mentale Prozesse, Entscheidungsgründe und mentale Modelle zu erheben. Es wird empfohlen, in einem ersten Schritt anhand von miteinander verbundenen Nachfolgeuntersuchungen die Grundmechanismen explorativ zu erforschen und die Robustheit der Prognosen aus Rahmentheorien zu evaluieren.

Literatur

- Ackermann, K. A., & Murphy, R. O. (2019). Explaining cooperative behavior in public goods games: How preferences and beliefs affect contribution levels. *Games*, 10(1), 15. doi: 10.3390/g10010015
- Allport, D. A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer & A. F. Sanders (Hrsg.), *Perspectives on perception and action*, (15, S. 395-419). London; New York, NY: Routledge
- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401), 464-477. doi: 10.2307/2234133
- Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics*, 116(1), 293-312. doi: 10.1162/003355301556419
- Alger, I., & Weibull, J. W. (2013). Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6), 2269-2302. doi: 10.3982/ECTA10637
- Alger, I., & Weibull, J. W. (2019). Evolutionary models of preference formation. *Annual Review of Economics*. doi: 10.1146/annurev-economics-080218-030255
- Axelrod, R. (1984). The evolution of cooperation. New York, NY: Basic Books.
- Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. A. Bower (Hrsg.), *Psychology of learning and motivation* (Band 8, S. 47-90). New York, NY: Academic Press.
- Betz, D., Esser, H., Biniossek, C. (2020) *Dataset Feedback, Punishment, and Framing in 30 Rounds Public Goods Experiments*. doi: 10.23663/x2648
- Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2), 191-208. doi: 10.1002/bdm.621
- Bigoni, M., & Suetens, S. (2012). Feedback and dynamics in public good experiments. *Journal of Economic Behavior & Organization*, 82(1), 86-95. doi: 10.1016/j.jebo.2011.12.013
- Biniossek, C. (2019). Prosoziales Verhalten in großen Gruppen. Köln: GESIS-Schriftenreihe (Band 21).
- Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1), 166-193. doi: 10.1257/aer.90.1.166
- Brosig-Koch, J., Riechmann, T., & Weimann, J. (2007). *Selfish in the end? An investigation of consistency and stability of individual behavior* (FEMM Working Papers Nr. 05). Magdeburg: Otto-von-Guericke Universität.
- Chaudhuri, A. (2018). Belief Heterogeneity and the Restart Effect in a Public Goods Game. *Games*, 9(4), 96. doi: 10.3390/g9040096
- Clark, K., & Sefton, M. (2001). Repetition and signalling: experimental evidence from games with efficient equilibria. *Economics Letters*, 70(3), 357-362.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. doi: 10.1037/0033-2909.112.1.155
- Croson, R.T. (2001). Feedback in voluntary contribution mechanisms: an experiment in team production. *Research in Experimental Economics*, 85-97.
- Engelkamp, J., & Cohen, R. L. (1991). Current issues in memory of action events. *Psychological Research*, 53(3), 175-182.

- Erev, I., & Haruvy, E. (2015). Learning and the economics of small decisions. In J.H. Kagel, & A.E. Roth (Hrsg.), *The handbook of experimental economics* Band 2, S.638-700). Princeton, NJ: Princeton University Press.
- Esser, H. (1996). Die Definition der Situation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48(1): 1-34.
- Esser, H. (2001a). Soziologie. Spezielle Grundlagen. Band 1: Situationslogik und Handeln. Frankfurt a. M./ New York: Campus..
- Esser, H. (2001b). Soziologie. Spezielle Grundlagen. Band 6: Sinn und Kultur. Frankfurt a. M./ New York: Campus..
- Esser, H. (2002a). In guten wie in schlechten Tagen? Das Framing der Ehe und das Risiko zur Scheidung. Eine Anwendung und ein Test des Modells der Frame-Selektion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54: 27-63.
- Esser, H. (2002b). Ehekrise: Das (Re-) Framing der Ehe und der Anstieg der Scheidungsraten. *Zeitschrift für Soziologie*, 31(6), 472-496.
- Esser, H. (2003). Der Sinn der Modelle. Antwort auf Götz Rohwer. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 55, 359. doi:10.1007/s11577-003-0055-z
- Esser, H. (2010a). Das Modell der Frame-Selektion. Eine allgemeine Handlungstheorie für die Sozialwissenschaften. Soziologische Theorie kontrovers. *Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 50, 45-61.
- Esser, H. (2010b). Sinn, Kultur, Verstehen und das Modell der soziologischen Erklärung. In M. Wohlrab-Sahr (Hrsg.), *Kulturosoziologie* (S. 309-335). VS Verlag für Sozialwissenschaften.
- Esser, H. (2017). *Rational Choice or Framing? RCT and MFS as Two Approaches to Explain the Patterns in the Fehr-Gächter-Experiments on Cooperation and Punishment in the Contribution to Public Goods*. Unveröffentlichtes Manuskript.
- Esser, H. (2018). Sanktionen, Reziprozität und die symbolische Konstruktion einer Kooperations-„Gemeinschaft“. *Zeitschrift für Soziologie*, 47(1), 8-28. doi: 10.1515/zfsoz-2018-1001
- Esser, H. (2019). Rational Choice oder Framing?. In J. Mayerl, T. Krause, A. Wahl, M. Wuketich (Hrsg.), *Einstellungen und Verhalten in der empirischen Sozialforschung*. Wiesbaden: Springer VS. doi: 10.1007/978-3-658-16348-8_3
- Esser, H., & Kroneberg, C. (2015). An integrative theory of action: the model of frame selection. In E. J. Lawler, S.R. Thye, & J. Yoon (Hrsg.), *Order at the Edge of Chaos: Social Psychology and the Problem of Social Order* (S. 63-85). New York, NY: Cambridge University Press.
- Ezekiel, M. (1938). The cobweb theorem. *The Quarterly Journal of Economics*, 52(2), 255-280.
- Fallucchi, F., Luccasen, R. A., & Turocy, T. L. (2019). Identifying discrete behavioural types: a re-analysis of public goods game contributions by hierarchical clustering. *Journal of the Economic Science Association*, 5(2), 238-254. doi: 10.1007/s40881-018-0060-7
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. Zanna (Hrsg.), *Advances in experimental social psychology* (Band 23, S. 75-109). London, GB; San Diego, CA: Academic Press. doi: 10.1016/S0065-2601(08)60318-4

- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980–994. doi: 10.1257/aer.90.4.980
- Fehr, E., & Gintis, H. (2007). Human motivation and social cooperation: Experimental and analytical foundations. *Annu. Rev. Sociol.*, 33, 43–64. doi: 10.1146/annurev.soc.33.040406.131812
- Fehr, E., & Schmidt, K. M. (1999). A Theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114(3): 817–868. doi: 10.1162/003355399556151
- Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10(2): 171–178. doi: 10.1007/s10683-006-9159-4
- Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American economic review*, 100(1), 541–56. doi: 10.1257/aer.100.1.541
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics letters*, 71(3), 397–404.
- Frings, C., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Hommel, B., Kiesel, A., Kunde, W., Mayr, S., Moeller, B., Möller, M., Pfister, R., & Philipp, A. (2020a). Merkmalsintegration und Abruf als wichtige Prozesse der Handlungssteuerung – eine Paradigmen-übergreifende Perspektive. *Psychologische Rundschau*, 71, 1–14. doi: 10.1026/0033-3042/a000423
- Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Kiesel, A., Kunde, W., Mayr, S. Moeller, B., Möller, M., Pfister, R., & Philipp, A. (2020b). Binding and Retrieval in Action Control (BRAC). *Trends in Cognitive Sciences*, 1–13. doi: 10.1016/j.tics.2020.02.004.
- Frith, C. D., & Haggard, P. (2018). Volition and the brain—revisiting a classic experimental study. *TRENDS in Neurosciences*, 41(7), 405–407. doi: 10.1016/j.tins.2018.04.009
- Gegenfurtner, K., Bremmer, F., Fiehler, K., Henriques, D., & Krauzlis, R. (2010). *Vision Research special issue on “Perception and action”*. doi: 10.1016/j.visres.2010.11.001
- Glassmann, U., & Sauermann, J. (2011). Entscheidungskosten und Gemeinwohleffekte demokratischer Abstimmungsregeln—eine experimentelle Untersuchung. *Politische Vierteljahresschrift*, 373–398. doi: 10.5771/0032-3470-2011-3-373
- Gräsel, C., & Gniewosz, B. (2015). Überblick Lehr-Lernforschung. In H. Reinders, H. Ditton, C. Gräsel, B. Gniewosz (Hrsg.), *Empirische Bildungsforschung* (S. 15–20). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Greiner, Ben. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association* 1(1): 114–125. doi: 10.1007/s40881-015-0004-4.
- Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental brain research*, 126(1), 128–133. doi: 10.1007/s002210050722
- Hanau, A. (1928). Die Prognose der Schweinepreise In: *Vierteljahreshefte zur Konjunkturforschung*, Verlag Reimar Hobbing: Berlin.
- Hoffmann, J. (2013). *Das aktive Gedächtnis: Psychologische Experimente und Theorien zur menschlichen Gedächtnistätigkeit*. Berlin; Heidelberg; New York, NY: Springer.

- Hommel, B. (1999). Temporäre Strukturbildung in Wahrnehmung und Handlung. In W. Hacker & M. Rinck (Hrsg.), *Bericht über den 41. Kongreß der Deutschen Gesellschaft für Psychologie in Dresden 1998*. Lengerich: Pabst.
- Hommel, B., Lippelt, D. P., Gurbuz, E., & Pfister, R. (2017). Contributions of expected sensory and affective action effects to action selection and performance: Evidence from forced-and free-choice tasks. *Psychonomic Bulletin & Review*, 24(3), 821-827. doi: 10.3758/s13423-016-1139-x
- James, W. (2007). *The principles of psychology* (Band 2). New York, NY: Cosimo. (Originalarbeit veröffentlicht 1890).
- Janczyk, M. (2016). Die Rolle von Handlungszielen bei der Entstehung von Doppelaufgabenkosten. *Psychologische Rundschau*. doi: 10.1026/0033-3042/a000324
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, 24(2), 175-219.
- Kandel, E. R. (2007). *In search of memory: The emergence of a new science of mind*. New York, NY: WW Norton & Company.
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (Hrsg.). (2013). *Principles of neural science* (5. Aufl.). New York, NY: McGraw-Hill.
- Koch, I., & Kunde, W. (2002). Verbal response-effect compatibility. *Memory & Cognition*, 30(8), 1297-1303. doi: 10.3758/BF03213411
- Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 284(1), 1-17. doi: 10.1007/BF00412364
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature reviews neuroscience*, 2(9), 635-642.
- Kroneberg, C. (2005). Die Definition der Situation und die variable Rationalität der Akteure: Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie*, 34(5), 344-363.
- Kroneberg, C. (2011). Das Modell der Frame-Selektion (MFS). In *Die Erklärung sozialen Handelns: Grundlagen und Anwendung einer integrativen Theorie* (S. 119-164). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kroneberg, C. (2014). Frames, scripts, and variable rationality: An integrative theory of action. In G. Manzo (Hrsg.), *Analytical sociology: Actions and networks* (S. 97-123). Chichester, GB: Wiley
- Kunde, W. (2006). Antezedente Effektrepräsentationen in der Verhaltenssteuerung. *Psychologische Rundschau*, 57(1), 34-42. doi: 10.1026/0033-3042.57.1.34
- Ledyard, J. O. (1995). Public Goods: A Survey of Experimental Research. In J. H. Kagel & A. E. Roth (Hrsg.), *The Handbook of Experimental Research* (S. 111-194). Princeton, NJ: Princeton University Press.
- Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and social psychology bulletin*, 30(9), 1175-1185. doi: 10.1177/0146167204264004

- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act. *Brain*, 106(3), 623-642. doi: 10.1093/brain/106.3.623
- Liebertson, S. (1994). More on the uneasy case for using Mill-type methods in small-N comparative studies. *Social forces*, 72(4), 1225-1237.
- Lijphart, A. (1971). Comparative politics and the comparative method. *American political science review*, 65(3), 682-693.
- Miettinen, T., Kosfeld, M., Fehr, E., & Weibull, J. W. (2020). Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions. *Journal of Economic Behavior & Organization*, 173, 1-25. doi: 10.1016/j.jebo.2020.02.018
- Mill, J. S. (1843). A system of logic, ratiocinative and inductive, being a connected view of the principles of evidence, and the methods of scientific investigation (Band 1). London: John W. Parker. Abgerufen von https://archive.org/details/bub_gb_y4MEAAAQAAJ/page/n1/mode/2up
- Mitzkewitz, M. & Nagel, R. (1993). Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory*, 22(2), 171-198.
- Münch, R. (1980). Über Parsons zu Weber: Von der Theorie der Rationalisierung zur Theorie der Interpenetration. *Zeitschrift für Soziologie*, 9(1), 18-53. doi: 10.1515/zfsoz-1980-0102
- Nann, M., Cohen, L. G., Deecke, L., & Soekadar, S. R. (2019). To jump or not to jump-The Bereitschaftspotential required to jump into 192-meter abyss. *Scientific reports*, 9(1), 1-9. doi: 10.1038/s41598-018-38447-w
- Nax, H. H., Burton-Chellew, M. N., West, S. A., & Young, H. P. (2016). Learning in a black box. *Journal of Economic Behavior & Organization*, 127, 1-15. doi: 10.1016/j.jebo.2016.04.006
- Neugebauer, T., Perote, J., Schmidt, U., & Loos, M. (2009). Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *Journal of Economic Psychology*, 30(1), 52-60.
- Nikiforakis, N. (2010). Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior*, 68(2), 689-702.
- Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Hrsg.), *Relationships between perception and action* (S. 167-201). Berlin; Heidelberg: Springer. doi: 10.1007/978-3-642-75348-0
- Prinz, W. (1997). Perception and action planning. *European journal of cognitive psychology*, 9(2), 129-154. doi: 10.1080/713752551
- Samak, A., & Sheremeta, R. (2013). Visibility of Contributors and Cost of Information: An Experiment on Public Goods. (MPRA Paper No. 46779). Munich Personal RePEc Archive, Munich. <https://mpra.ub.uni-muenchen.de/46779/>
- Sauermann, J. (2018). Do individuals value distributional fairness? How inequality affects majority decisions. *Political Behavior*, 40(4), 809-829. doi: 10.1007/s11109-017-9424-6
- Sauermann, J., & Kaiser, A. (2010). Taking Others into Account: Self-Interest and Fairness in Majority Decision Making. *American Journal of Political Science*, 54(3), 667-685. doi: 10.1111/j.1540-5907.2010.00453.x

- Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of vision*, 11(9), 1-30. doi: 10.1167/11.5.9
- Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In H. Sauermann (Hrsg.), *Beiträge zur experimentellen Wirtschaftsforschung* (S. 136-168). Tübingen: J.C.B. Mohr (Paul Siebeck).
- Stocké, V. (2002). *Framing und Rationalität: die Bedeutung der Informationsdarstellung für das Entscheidungsverhalten*. München: Oldenbourg.
- Thöni, C., & Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters*, 171, 37-40. doi: 10.1016/j.econlet.2018.06.022
- Varian, H. R. (1994). Sequential Provision of Public Goods. *Journal of Public Economics*, 53(2), 165-186.
- Vesterlund, L. (2006). Why do people give. *The nonprofit sector: A research handbook*, 2, 168-190.
- Weber, R. A. (2003). 'Learning' with no feedback in a competitive guessing game. *Games and Economic Behavior*, 44(1), 134-144.
- Weimann, J. (1994). Individual behaviour in a free riding experiment. *Journal of Public Economics*, 54(2), 185-200.
- Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6(3), 299-310. doi: 10.1023/A:1026277420119

Anhang I: Questionnaire

Item	Text in Questionnaire
<i>otherdecision_1</i>	Ich hätte mir gut vorstellen können, mich anders zu entscheiden.
<i>emotion_1</i>	Ich habe die Situation als angenehm empfunden.
<i>outcomemax_1</i>	Ich wollte eine möglichst hohe Geldauszahlung haben.
<i>fair_1</i>	Ich habe versucht, mich fair zu verhalten.
<i>reciprok_1</i>	Ich wollte nicht mehr beitragen als die anderen, weil ich nicht der Dumme sein wollte.
<i>secondorder_1</i>	Mich so zu verhalten, wie es auch andere tun würden, war mir wichtig.
<i>change_2</i>	In welcher Periode hat sich Ihre Wahrnehmung der Situation geändert?
<i>otherdecision_2</i>	Ich hätte mir gut vorstellen können, mich anders zu entscheiden.
<i>emotion_2</i>	Ich habe die Situation als angenehm empfunden.
<i>outcomemax_2</i>	Ich wollte eine möglichst hohe Geldauszahlung haben.
<i>fair_2</i>	Ich habe versucht, mich fair zu verhalten
<i>reciprok_2</i>	Ich wollte nicht mehr beitragen als die anderen, weil ich nicht der Dumme sein wollte.
<i>secondorder_2</i>	Mich so zu verhalten, wie es auch andere tun würden, war mir wichtig.
<i>change_reason_perception_3</i>	Uns interessiert nun, ob sich im weiteren Verlauf der drei Experimente nochmals grundlegende Änderungen der Situation für Sie ergeben haben. Bitte beschreiben Sie kurz alle wesentlichen weiteren Änderungen, Ihre jeweilige Wahrnehmung der Situation, Ihre Überlegungen und was Ihr jeweils wichtigster Entscheidungsgrund war.
<i>age</i>	Wie alt sind Sie?
<i>gender</i>	Was ist Ihr Geschlecht?
<i>experience</i>	Wie häufig haben Sie schon an Experimenten teilgenommen?
<i>student</i>	Studieren Sie?
<i>semester</i>	Wenn ja, in welchem Semester?
<i>discipline</i>	Was studieren Sie?

4 Fixation durations – Why are they so highly variable?¹

Niels Galley¹, Dirk Betz^{2,3,*} and Claudia Biniossek^{2,3}

¹ University Cologne, Center for Anatomy, Cologne, Germany

² GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany.

³ University of Cologne, Institute of Sociology and Social Psychology, Cologne, Germany.

Keywords: eye movement, visual processes, express fixation, fixation duration, saccades, mental processes, electrooculogram (EOG)

Abstract

Fixation duration is the time during which the eyes rest on an object in the surroundings. Fixation lasts approximately 250 milliseconds (ms) but often is shorter or longer. It is usually defined as the time between the end of one saccade and the beginning of the next one. Usually, we fixate on something in our surroundings consciously. However, fixation duration can vary between several milliseconds, which is too short for conscious control, and several seconds. The reasons for this large variation are not obvious. In this book chapter we look at eye movement data collected with an electrooculogram (EOG) from a large population. A total of up to 1131 subjects performed several tasks, for example looking at a picture, tracking a moving spot, solving an intelligence test (Raven matrices) or closing the eyes, a situation in which saccades do not pause. Particularly, we reinforce the view of fixations as a mental or cognitive process that can be separated into several partitions. For example, in the past the so-called *express fixations* in the range of less than 100 ms to approximately 150 ms were assigned to a special group of short fixations and were controversial. These shorter fixation durations as well as ones that last for seconds

1 Postprint: Galley, N., Betz, D., & Biniossek, C. (2015). Fixation durations: Why are they so highly variable. In T. Heinen (Hrsg.), *Advances in Visual Perception Research* (S. 83-106). New York, NY: Nova Science.

* Corresponding author: GESIS – Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Cologne, Germany. Email: dirk.betz@gesis.org.

do occur, and these *very short* or *very long* fixation durations have seldom attracted the interest of researchers. Visual processes and mental processes can be more or less connected. In this book chapter we try to determine some reasons for the obvious high variability of fixation durations, which occur in both the closely and loosely connected modes of the visual-motor and the cognitive system.

Introduction – Why Are Fixation Durations So Different?

In the EOG of a child (see Figure 1) twelve saccades and subsequent fixations can be seen. The child was looking at a table of Raven Matrices (Raven, Court, & Raven, 1978) and trying to figure out which of the six alternatives shown at the bottom correctly filled in the gap at the top.

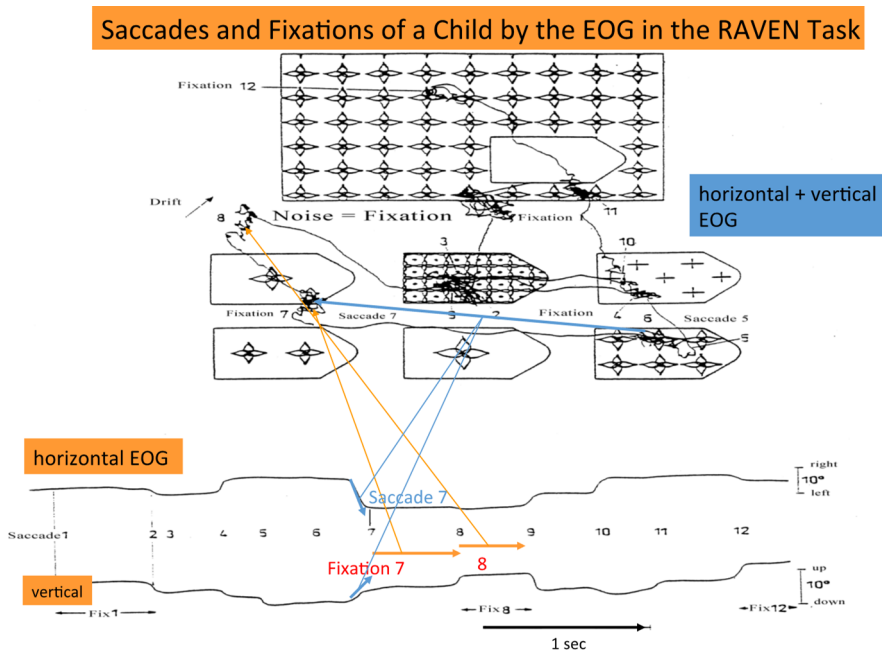


Figure 1 In the lower field is the horizontal and the vertical EOG of a child who looked at the Raven table (in the upper field) and tried to decide which of the six alternatives correctly filled the gap in the upper figure. The two-dimensional EOG is spatially overlaid on the figure, which represents fixations as tangles and saccades as straight lines.

Fixation 1 lasts 800 ms, during which the eyes rest on the gap at the bottom of the pattern, which is to be filled with the correct option. Saccade 2 goes down to the second option, and the eyes only rest there for approximately 100 ms. Then, saccade 3 goes horizontally to the third alternative with a fixation of approximately 400 ms.

The eyes of the child move further along the lower pattern with the six alternatives to fill the gap until saccade 11 again reaches the upper pattern. After saccade 12 and its accompanying fixation duration 12, which had a duration of 300 ms, the child gave the correct answer by pointing to ‘alternative 6’. Interestingly the eyes hadn’t rested on alternative 5, and one can speculate if the child had overlooked this choice or excluded this false choice using peripheral vision when the gaze rested at nearby alternatives 6, 2, 4 or 1.

As one can see in Figure 2, most fixations of adults on a *Raven* task designed for adults are shorter. The *median* for the easier to solve tables was 230 ms, and the median for the more difficult tables was 260 ms.

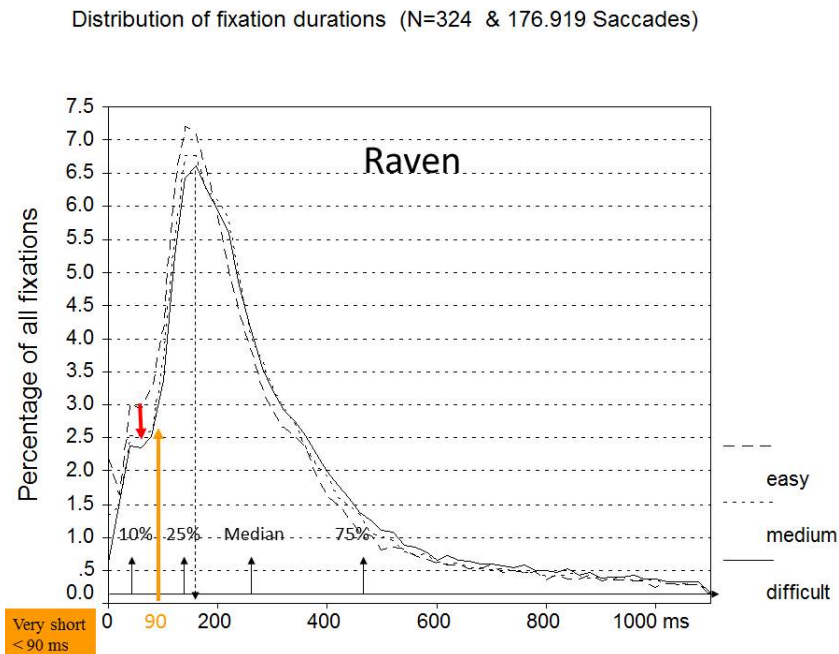


Figure 2 Three frequency curves for fixation durations (in milliseconds = ms) for a subpopulation of 324 students who solved ten *Raven* tables of increasing difficulty. There is a small shift to longer fixation durations when looking at the more difficult *Raven* tables

The skewed distribution curves show that most of the adult fixations are shorter than those of the child: the *modal* value is 190 ms. However, because there are many long to very long fixations, the median (approximately 250 ms), which is the most used descriptive value of the central tendency, was higher than the most frequent value, which is also known as the modal value. Thus, the modal value seems to be the most obvious choice to describe the distribution curves, instead of the median, which is the more common choice in the literature. It is unnecessary to emphasize that the *mean* value is not appropriate to characterize values of fixation duration, because

this description of the central tendency is very dependent on some very high values, which always occur. Nevertheless, the mean is often used in the literature (Schweiger, 2003), most likely after the range of too short and too long fixations are eliminated.

It seems obvious from these figures that the reasons for the variability of fixation duration are not evident at first glance but must be deduced from factors that have a measurable effect on duration. Figure 2 shows three distribution curves for fixation duration: one comes from the easy to solve *Raven* tables, one comes from the difficult to solve tables and one comes from those of moderate difficulty. The three curves are *shifted a little bit to the right for longer durations*. This is an indication of *one small factor*, which is responsible for different fixation duration values: *the more difficult the task for the subject, the longer the eyes remain on the fixated object*.

Another feature of these three curves seems noteworthy: in the area of very short fixations, roughly less than 90 ms, there are fewer values if the task is more difficult (marked by the red arrow). This is a typical effect of the *inhibition of irrelevant behavior*. More mental effort indicates longer fixations not only by inhibiting the release of the next saccade but also by specifically impeding very short fixations. This can be an effect of *concentration*, which seems to induce a shift to more conscious, late operations and less spontaneous, early ones. The difficulty of the task and the induced mental effort of the subject seem therefore two small factors among many other probable factors that influence fixation duration.

In Figure 1, the child's eyes rest between 100 and approximately 800 ms. In contrast, if a robotic device had scanned this pattern, there would be a very uniform rate. Perhaps some of the child's *internal mental processes* need different lengths of time before the eyes are allowed to move to the next position. However, can there be a *thinking process* that lasts only 100 ms? For example, in such a manner: '*Alternative 3 cannot be the right answer*'. Surely, 'No' should be the answer to this question. This seems to be too short a time to *recognize* the visual object, *to decide something* and then to *initiate another look*. One has to mention that with each shift of the gaze the subject has two tasks: The first is to analyze the stimuli in the central vision and the second is to decide where in the periphery to look next. A new saccade is presumptively not generated until the object(s) falling on the fovea had been sufficiently analyzed. Then, one has to select which object in the visual scene should be looked at next. Some basic analysis needs to be carried out in the near or far periphery to allow for a desirable selection. Thus, very short fixations of <90 ms seem to have no opportunity to be controlled by conscious cognitive operations. Longer fixations in the range of 'express fixations' - as the range between 90 and 150 ms is called - have very little time for cognitive operations, so they will presumptively go to *expected locations* or otherwise foreseen locations. However, what sort of control processes can be assumed for this very short duration behavior of 80 or 70 ms? We will return to this point later on. Now it seems necessary to explain how to measure the exact time of such short fixation durations.

How to Measure Fixations?

To measure fixation duration, one has to identify the *saccades* that flank fixation and bring the eyes (and then the visual centers in the brain) into powerful contact with a new object because the fovea is the retinal place where this new object is projected. The fovea has the highest spatial resolution, and its central representation has preferential access to consciousness. For example, it is seldom consciously recognized that most objects in our peripheral vision are doubles.

Saccades are movements of the eyes with very high velocities that are flanked by rests of the eyes, so the human inspector has no difficulties identifying saccades even in a noisy signal such as the electrooculogram (EOG), because these rests are seen as plateaus before and after abrupt changes. We used EOG in our research on fixations. As seen in Figure 3, saccades in the EOG look like large stairs if they were responses (green beams) to the target jumps or like small stairs between jumps.

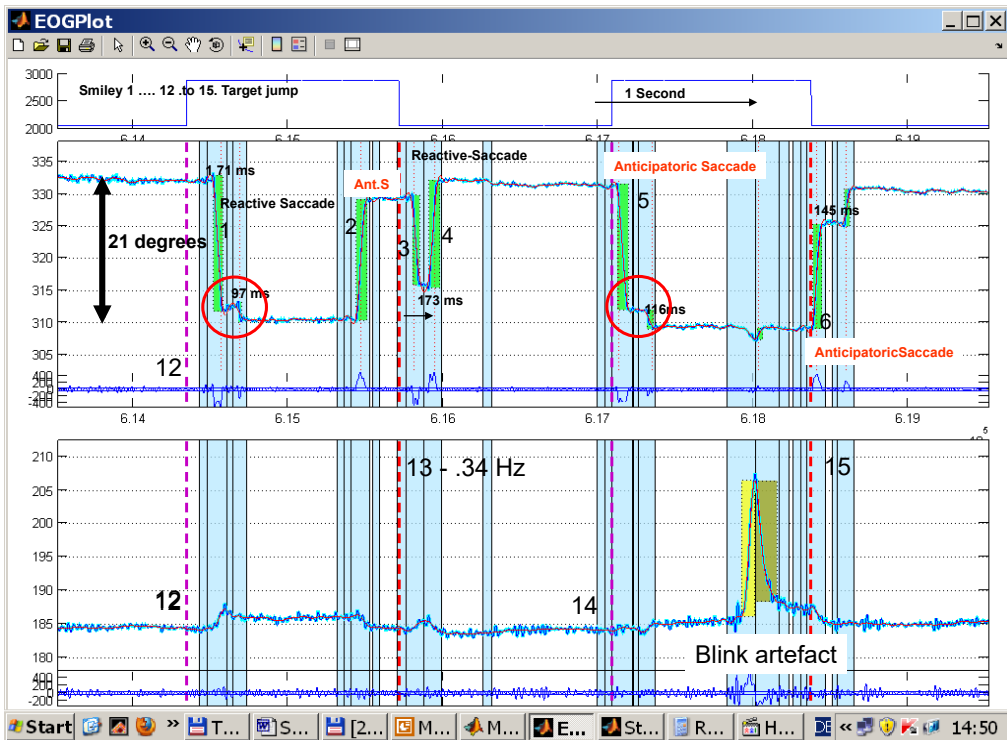


Figure 3 A screenshot of our computer program, EOGUI, which shows the horizontal EOG (first row) and the vertical EOG (second row) for a subject who followed a jumping target spot. The six large saccades that were identified are shown as green beams. They are gaze responses of the subject to target jumps number 12, 13, 14 and 15 in the sequence of 112 jumps with decreasing intervals. In the areas of the EOG that are circled in red, very small saccades can be detected, which were not identified by the program although the human inspector clearly traced these small steps as saccades.

In the task, *Jumping Point*, the subject had to follow a jumping target point on a screen with his gaze “as closely as possible”. The target point jumped over a distance of 21 degrees (in this example) with continuously diminishing intervals. At the beginning, the frequency of the jumping was slow (0.2 Hz), but it sped up to 1.4 Hz, which is very demanding. Figure 3 is a detail of the horizontal (upper row) and vertical EOG channels (lower row) in which a blink artefact can be seen. Starting from jump number 12 the first response saccade (1) occurred 171 ms later. The next saccade (2) occurred approximately 250 ms before jump number 13 and is thus an anticipatory saccade. Very shortly after the jump a return saccade (3) can be seen. Very shortly afterwards another response saccade (4) occurred. Therefore, we had to decide if saccade 2 or 4 was the response saccade, because more than one response saccade for one target jump is not appropriate. We decided that the large saccade in the correct direction that came closest to the time of the target jump should be counted as the response saccade and all other saccades should not. In this case saccade 4 is the winner, because it is nearer to the time of the target jump than saccade 2.

A green beam overlaying the EOG signifies that our computer program EOGUI has identified a saccade at that time. Here is a URL where people can download our program: <http://www.mathworks.com/matlabcentral/fileexchange/32493>.

The program is and has to be *conservative*: It is better for some very small saccades to be missed than for some noise to be identified as saccades. There are some consequences of using our machine to identify fixation durations. Every time a small saccade is missed, the fixation duration that is registered is too long. However, as one can see in figure 4, saccades aren't normally very small. If they were 2 degrees or larger, they were mostly well identified by the program. The percentage of saccades not identified by the program is 5 to 6 percent. By using the median of many fixations per subject and per task, the presence of some artificial, overly long fixations does not matter.

What Can the EOG (Not) Tell Us about Fixations and Eye Movements?

We have mentioned that noise in the electrooculogram can prevent saccades smaller than 1 to 2 degrees from being identified by our computer program. However, there is another weakness of EOG: as you can see in Figure 1, the EOG is superimposed over signal drift. In concrete terms, the signal is not positionally stable over a long time period. It has a very slow inherent shift, which makes determining exact position information of the eyes impossible. Thus, the sometimes very important information some users hope to obtain from eye movement registration, namely, *where* the subject is looking, is not simply available from EOG. Therefore, in a comprehensive guide to methods and measures (Holmqvist, Nyström Andersson, Dewhurst, Jarodzka & Van de Weijer, 2011) for eye movement registration systems, EOG is only mentioned as a low-cost method that is not useful for research. This seems to be a too strict verdict, but it is worth asking what research *should not be conducted* with EOG:

- All research which needs exact information about the position of the gaze in relation to the target.

- All research on saccades of very small amplitude, which are called microsaccades. The question of how many microsaccades occur is relevant. Our computer program identified overly long fixation durations if a microsaccade was missed. However, fortunately microsaccades aren't very frequent (see below), and the few occurrences were made less important by using the median of the fixation duration.

There are at least three advantages of EOG over other systems of eye movement registration:

- 1) EOG and thus saccades, fixations and blink artefacts can be collected when *the eyes are closed*. The Eyes Closed Condition occurs not only during sleep and can serve as an interesting control condition for tasks with open eyes. Saccades are still present during very light sleep as well as during so-called REM sleep, which is *rapid eye movement* sleep while dreaming. The name REM was chosen by Aserinsky and Kleitman (1953), who were unaware that these rapid eye movements were saccades. However, rapid eye movements or saccades, as we now know, are absent in deep slow wave sleep, when the EEG is dominated by delta waves.
- 2) In the vertical EOG channel, there is an obvious *blink* artefact (see Figure 3), which enabled us to detect all blinks and to analyze the circumstances of their occurrence. Blinks are short time interruptions of vision due to the covering of the pupil by the lid. A blink lasts for approximately 200 ms. Blink rates are very variable between people and signify attention or variations in vigilance as well as characteristics of the subject's arousal level.
- 3) As an electrical signal, EOG has good time resolution. Thus it is possible to obtain very exact fixation durations on a millisecond basis, which is crucial for our present topic.

How often Do Microsaccades Occur, and What Is the Function of These Very Small Saccades?

Because very small saccades, called microsaccades, are sometimes missed by computer identification one might wonder how often or how seldom these small saccades occur. In two comprehensive reports about microsaccades (Otero-Millan, Troncoso, Macknik, S.L., Serrano-Pedraza I. & Martinez-Conde, 2008; Rolfs, 2009) the authors wrote that microsaccades have a mean amplitude of half a degree and that their occurrence is a function of the task. Microsaccades become more frequent the longer a fixation duration lasts. However, they are rare during fixations shorter than 250 ms. Microsaccades occur unconsciously but can be inhibited by will. Rolfs (2009) listed seven functions for these very small saccades, of which the *reduction of perceptual fading*, the *continuity of perception* and the *control of fixation position* seem to be the most important ones.

In our *used standard tasks* histogram, curves of saccadic amplitudes show notable peaks at 2, 3 or 4 degrees and declines at smaller amplitudes (see Figure 4). The exceptions to this rule were the tasks *Eyes Closed* and *Interview*, during which small saccades, i.e., microsaccades, were frequent. However, during these tasks the subjects had no visual

task. When the subjects performed *Picture Inspection* or followed a wandering sinusoidal point with their gaze (*Sinus*), solved the *Raven* task, or followed the *Jumping Point* they seldom showed very small saccades. Because we used the medians as the central tendency for fixation duration, the fact that the computer program overlooked some microsaccades had a negligible effect on fixation duration.

Saccade amplitudes differ in relation to the task, but amplitudes larger than 10 degrees are seldom seen with the exception of the *Jumping Point* task, in which the task demands a specific amplitude saccade to be performed by the subject. Usually, the target point jumped 12 degrees. The value was 21 degrees in only a few studies, as seen in Figure 3. The responses to the target jumps are usually large amplitude saccades, but small saccades with short latencies do occur, for example when correcting the landing position of the gaze. Correspondingly, the peak of the amplitudes in this task (see Figure 4) was at 11 degrees, mirroring the tendency of subjects to undershoot the target amplitude by 10 percent. The tendency to make small saccades to correct the undershoot is seen in the increase at 2 degrees for the curve of the jumping point shortly after the large response saccade. The function of other small saccades often remains obscure but may belong to the above-mentioned categories of microsaccades, to prevent the fading of the picture in central representation.

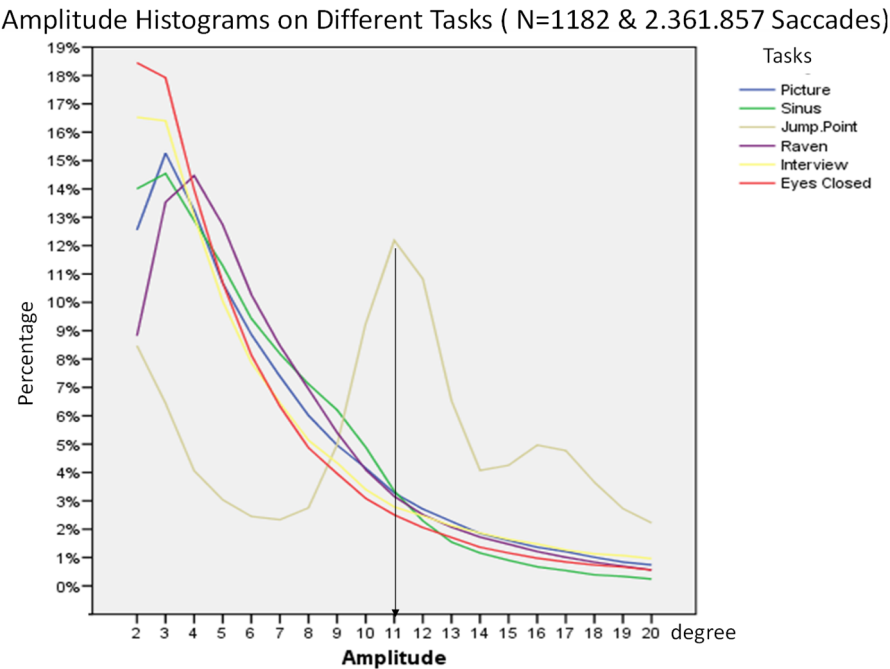


Figure 4 Distribution of histogram curves of saccadic amplitudes (in degrees) in our standard tasks: *Picture Inspection*, *Sinus*, *Jumping Point*, *Raven*, *Interview*, and *Eyes Closed*. In most tasks, amplitudes of 2, 3 or 4 degrees dominate, with the exception of *Jumping Point*, in which most subjects [but not all (see Figure 3)] had to follow a 12 degree jump of the target.

Our Standard Battery of Tasks

In our research on eye movements with several populations (children, adults, very old people, but mostly students) we used some tasks consistently in different studies. Some of the results are published (Galley 1984; Galley 1987a; Galley, 1987b; Galley and Fritz 1987; Galley 1993; Galley 1998; Galley and Galley 1999; Galley 2001; Galley 2010), but here we report analyses of pooled data under the main aspect *which factors modulate the fixation duration*.

In the task *Picture Inspection*, subjects looked at an illustration by Ali Mitgutsch in which many events occur. Subjects are supposed to remember these events because they will be asked about them later. *Picture Inspection* was usually the first task in our standard battery of tasks. The next two were *Sinus* and *Jumping Point*. Tasks where a moving spot was accelerated from a very slow 0.2 Hz to a fast 1.2 Hz (for children) or to 1.6 Hz for adults over 12 or 20 degrees were strenuous. In the task *Sinus*, the horizontal sinusoid's target spot wanders back and forth, following the same frequency scale as in the task *Jumping Point*, but the *pursuit eye movement system* is prominently involved in this task (Thier & Ilg, 2005), while only the saccadic eye movement system is involved in *Jumping Point*. Subjects followed the *sinus* target with prominently smooth pursuit eye movements, nevertheless, many saccades can be seen. Thus the rests of the eyes on the moving target are overlaid by abrupt changes of position from time to time. Usually these saccades aren't large, thus, the target does not leave the zone around the fovea from which the pursuit system controls the eyes. Increasing the velocity of the sinusoidal movement of the target increases the incidence of bigger saccades.

In most of our own studies cited above, these tasks were followed by the *Raven* task in a children's form (see Figure 1) or a corresponding adult form. Here subjects had to decide which of the alternatives was the correct one to fill the gap in the top figure.

The last two tasks were the *Interview* and *Eyes Closed* tasks. In the *Eyes Closed* task subjects were asked which details of the viewed picture they remembered, while in the task, *Interview*, subjects were questioned about which tasks they judged most or least difficult, most interesting, and so on. In these last two tasks there was no visual presentation that could guide the gaze. During *Eyes Closed*, only the remembered events of the Ali-Mitgutsch-Picture could guide the path of the eyes, and in *Interview* the outside environment was present but irrelevant.

In several studies devoted to different themes, most of our subjects completed these six standard tasks, which lasted only two minutes or less, successively. Only the *Raven* task lasted for several minutes. In every task subjects made hundreds or more fixations. For each fixation the computer identification captured and parameterized the following values: the beginning and end of each saccade in milliseconds as well as in spatial coordinates and the maximal velocity of the saccade. Additionally, all blinks, which sometimes mask the occurrence of a saccade, were identified by the algorithm and stored together with the values of the saccades on the hard disk. Pooling several studies for which our standard battery of tasks was used, we obtained a rather large number of subjects ranging from 4 to 90 years old and a huge amount of fixations.

Fixation as a Time-Consuming Process becomes Cognitive Controlled Not until after a while

We reported that sometimes fixation durations are very short, for example only 90 to 100 ms long. We wondered about these *express fixations*: How could visual information be processed so quickly by brain circuits? Visual information has to go through several stations in the retina, be transferred to the visual centers, and reach the oculomotor centers in the brainstem, and it must do all this in a few milliseconds. However, as one can see in Figure 5, a small fraction of fixation durations are *even shorter than 90 ms*. In this figure, the same data as in Figure 2 are analyzed and presented in a slightly different way.

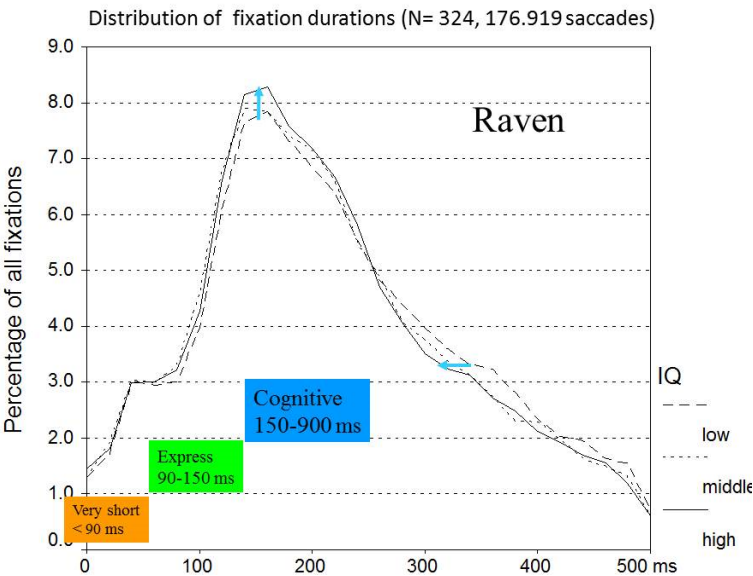


Figure 5 The same data are used here as in figure 2, but now the subject fixation durations during the *Raven* task are arranged in relation to three components of the subjects' IQ (intelligence scores). Additionally, only fixation durations shorter than 500 ms were considered. Curves of the more intelligent subjects were shifted to the left, i.e., to shorter values.

We labelled the shortest range of fixations as *very short* (1) fixations, the next range as *express* (2) and the longer fixations as the *cognitive* (3) range. This is programmatic, and we will present evidence to justify these categorizations and the borderlines between these categories. In comparison with figure 2, we cut off the tail at *very long* (4) fixations in figure 5, but we shall go back to this category later on. The data in figure 5 were arranged into three curves, according to the IQs of the subjects. In figure 2, the three histograms show a slight trend for longer fixations when subjects solved more difficult problems, i.e., a shift to the right. In figure 5 the rightward shift

is caused by the *lower competence of the subjects*. This corresponds nicely, because for people with lower intelligence the same problem seems to be more difficult. In other words, subjects with less competence *need more time for every fixation*. Thus, one factor that extends fixation duration seems to be the difficulty of the task or the lower competence of the subject.

The peculiarity of *express latencies* has attracted the attention of researchers due to the astonishingly very fast reaction of saccadic eye movements (less than 100 ms) in response to visual targets (Fischer & Weber, 1993). It has been proven that express reactions are truly visual *reactions* and not anticipatory behavior. However, there are to date open questions about what factors influence the time window of express reactions or express fixations. Thirty years ago, the typical way to obtain more or less express fixations in the laboratory was to have the subject fixate on a small spot, wait until a target appeared, and then look at the target. Usually, the latency for the unexpected target is approximately 200 ms, which is double that of express reaction time. But when the fixation stimulus disappears 200 ms before the target appears – a so called *gap paradigm* – express latencies of approximately 100 ms latency often occur. Later on, it was found that express fixations are often a rather small part of the normal reaction paradigm called the *overlap condition* and are seen in a multitude of normal visual tasks, such as reading (Radach, Heller & Inhoff, 1998) or searching (Sommer, 1994; McPeck, Skavenski & Nakayama, 2000).

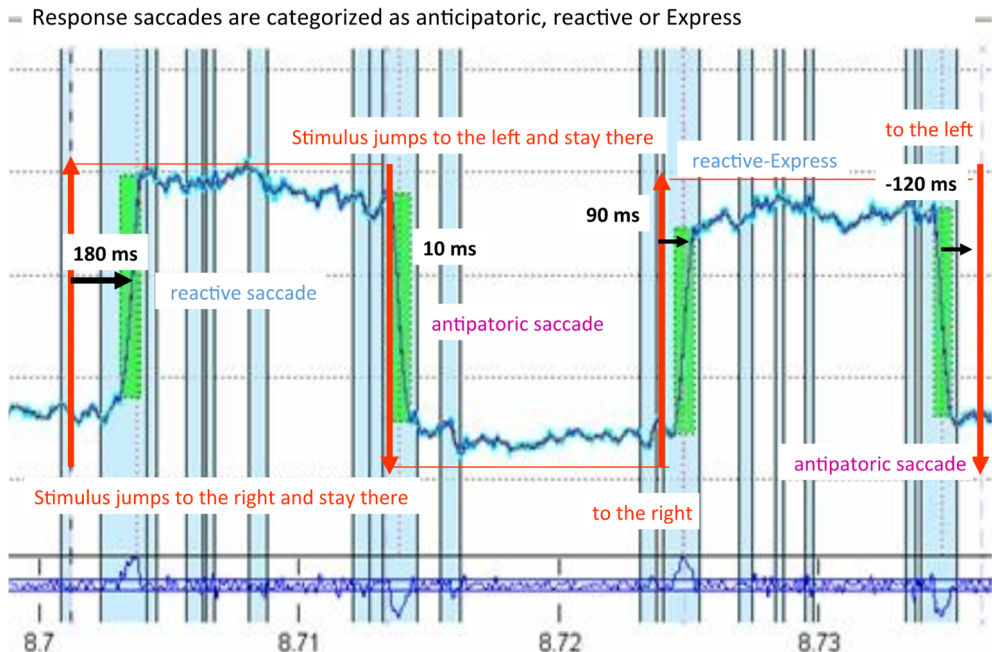


Figure 6 A section from the original registration of the horizontal EOG shows response saccades to target jumps as green beams. The subject's first response shown is a reaction with 180 ms latency after the jump. The next three are: an anticipatory response only 10 ms after the jump, which cannot be a reaction, an express reaction with 90 ms latency and then an anticipatory response, which occurs 120 ms before the target jumps.

In our task *Jumping Point* (see Figure 3 and 6) we learned about shortening response reaction time. At the beginning, subjects showed response saccade reaction times of approximately 200 ms. Later on, subjects tended to shorten their reaction times (Figure 6: 180 ms, 10 ms, 90 ms) or began to anticipate the jump (- 120 ms) when the intervals between the jumps became shorter. *Response saccades* are those that were in the same direction as the target jump, more than half of the amplitude of the target jump and closest in time to the jump.

The latencies of the responses do not indicate which times are reactive and which are anticipatory. A response 10 ms after the jump is obviously not a reaction but should be considered an anticipatory response. Responses 90 ms after the jump can be *express* reactions, and responses -120 ms before are in *anticipation*. A latency histogram of all response saccades from the large population of subjects who performed the task is shown in figure 7. It provides clues to help categorize the *anticipation*, *express* and *reaction* response modes.

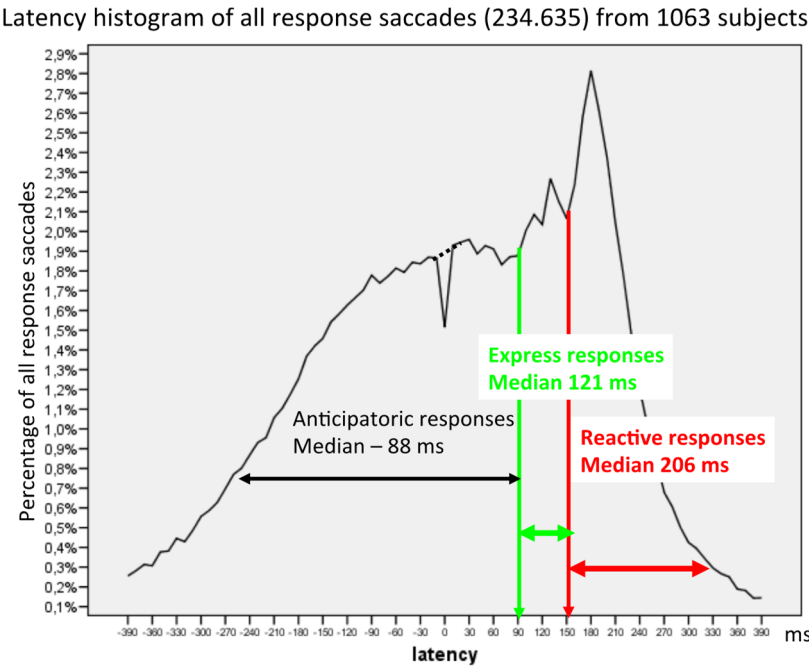


Figure 7 A latency histogram for all the response saccades from a large population ($N = 1063$) shows three sections or peaks. The first, labelled as *anticipatory*, ended approximately 90 ms after the target jump and had a median of - 88 ms ahead of the target jump. The second, labelled as *express*, ended at 150 ms and had a median of 121 ms. The third, called *reactive*, began at this time and had a median of 206 ms. The hollow at 0 ms is an artefact of the SPSS program that results from class construction around zero and is corrected by a dotted line.

One can see three sections or peaks in this histogram: a very broad peak for probable anticipatory responses and two small peaks at 122 ms and at 180 ms after the jump.

These curves led us to *define the express range* as the time window between 90 to 150 ms on a rather small sample (Galley 1990). After we have collected data for a large population, we can now assess whether this decision was correct. Evidence for fundamental differences between *anticipatory*, *express* and *reactive* response saccades comes from the different velocities and amplitudes of these response categories (see Figures 8 and 9).

Computing the standardized mean velocity of the response saccades shows different velocities for the values (see Figure 8): anticipatory response saccades had lower velocities than express or reactive ones over the whole range of jumping frequencies.

Response saccades differ in their velocities (N=1063 & 234.635 Saccades)

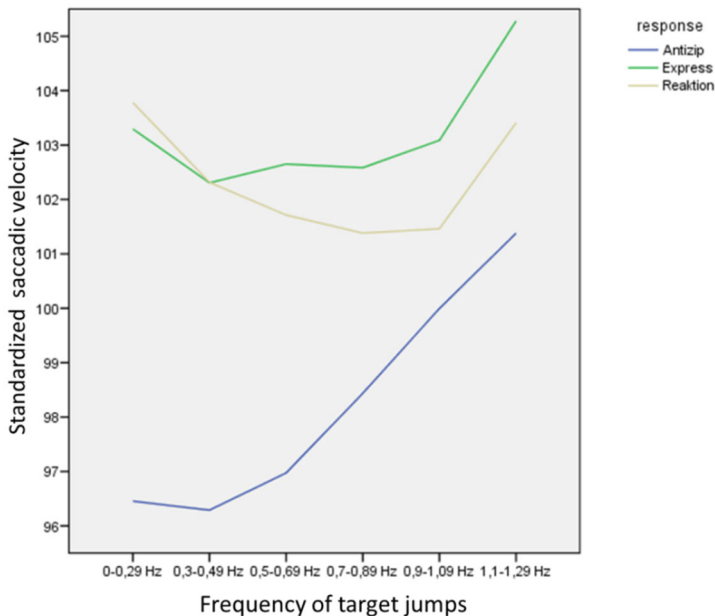


Figure 8 Standardized saccadic velocities in the three response categories (plotted as percentage of expected value) clearly differ. *Anticipatory* responses have lower velocities than *reactive* or *express* ones over the whole range of jumping frequencies.

Standardization of saccadic velocities is necessary because larger amplitude saccades have higher velocities, which is an undesirable effect.

Standardization computes the velocity of a specific saccade as the percentage of the value this velocity normally owes to this amplitude in a large population (Schleicher & Galley, 2005). For example, if a specific saccade with an amplitude of 11 degrees had a velocity of 200 degrees/sec and the normal value was 220 degrees/sec, then the standardized velocity is $200 \cdot 100 / 220 = 90.9$ percent of the expected value. By this standardization, saccadic velocity becomes a value of the momentary activation, which varies as a function of fatigue, motivation, effort and so on (Galley, 1998).

Analyses of the amplitudes of response saccades in *Jumping Point* additionally show that anticipatory responses tend to have smaller amplitudes than reactive or express responses for the first periods up to 0.6 Hz (see Figure 9).

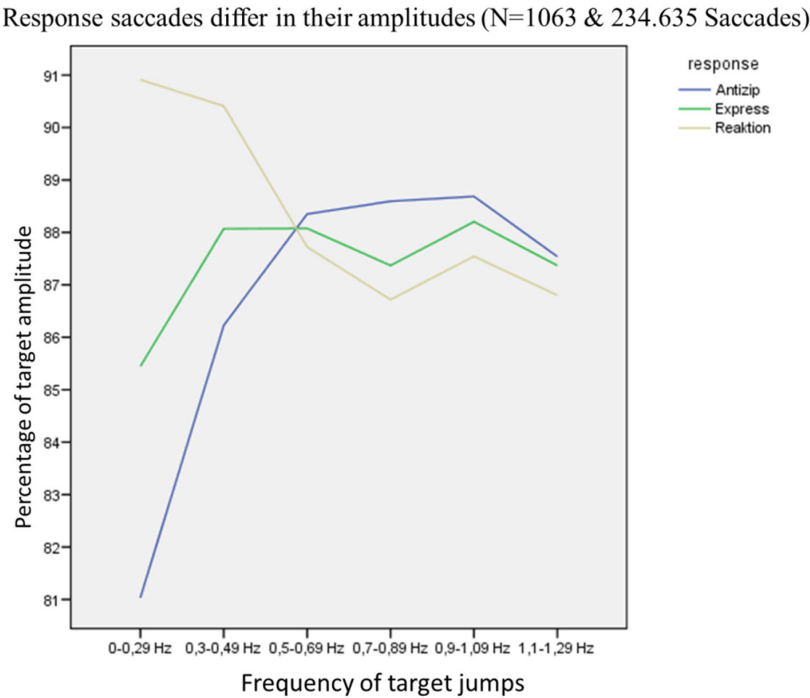


Figure 9 The amplitudes of response saccades with increasing target jump frequency in the different response categories are shown here. Anticipatory response amplitudes in the low frequency range from 0.2 and 0.4 Hz are clearly smaller than the express and reactive ones. Later on, when the jumping frequency increases, the amplitudes become more similar.

Using these different properties of anticipatory, express and reactive responses one can locate the borders of the categories (Galley, 1990) over the latency. Computing the standardized velocity of the response saccades over the latency of the response we found a characteristic increase roughly at the time when express responses began and a decrease when reactive responses began to occur (see Figure 10).

Histogram of saccadic velocities (234.635 response saccades) from 1063 subjects

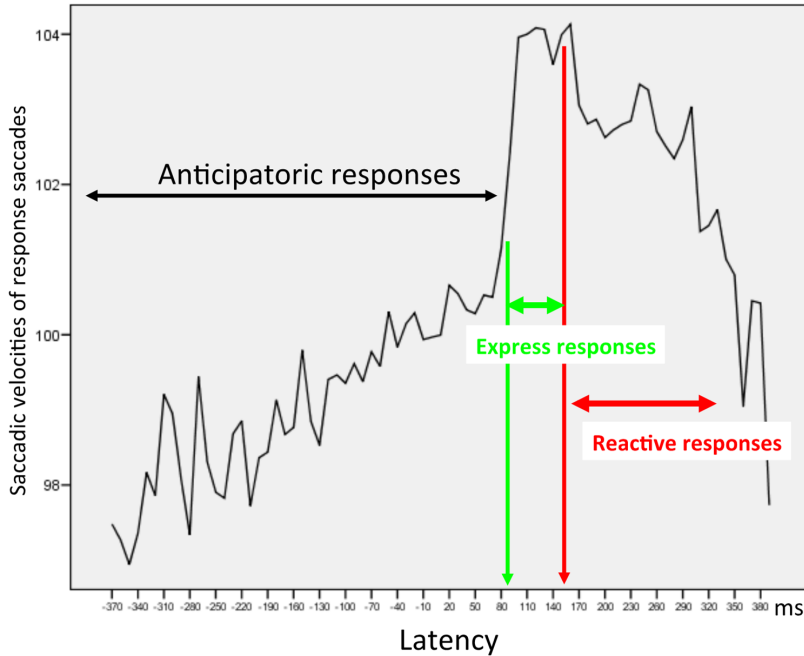


Figure 10 Response saccades show an abrupt increase in (standardized) velocity (plotted as percentage of expected value). Increase begins at 80 ms after the jump. Close to 170 ms one sees a downward trend.

The cause of the lower anticipatory saccade velocities is not known, but other research also confirms the lower velocities (Bronstein & Kennard, 1987; Smit & Van Gisbergen, 1989). Our guess is that that anticipatory behavior spared energy by not withholding the gaze while waiting for the appearance of the target (Galley, 1987b; Galley, 1989).

We have shown in Figure 9 that the amplitudes of anticipatory responses are smaller than reactive or express response saccades. Thus one can expect the amplitudes to show a profile over the latency of the response that reflects this behavior (see Figure 11).

Histogram of saccadic amplitudes (234.635 response saccades) from 1063 subjects

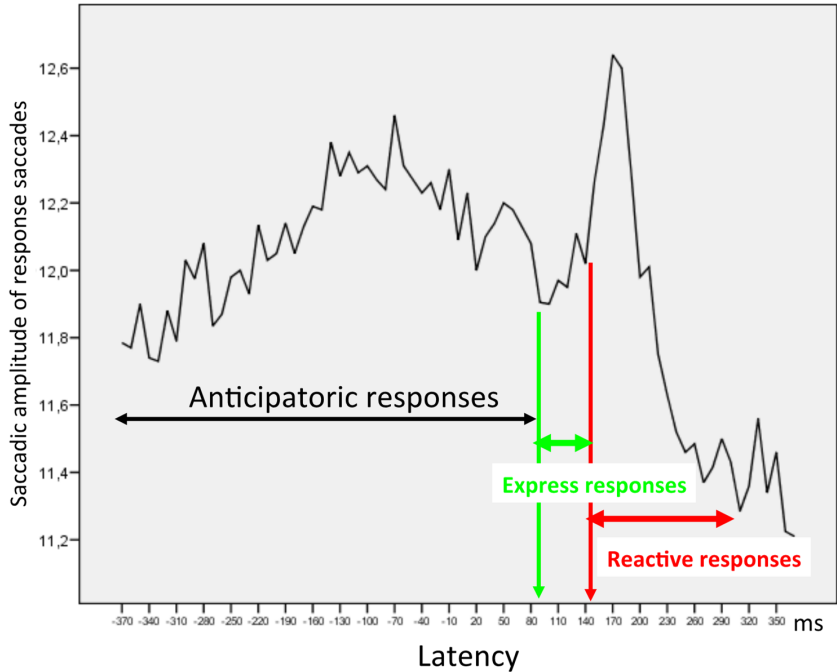


Figure 11 Response saccades show an abrupt increase in amplitude (in degrees) beginning roughly at 150 ms after the jump and a hole in the express range (90-150 ms).

Different saccade amplitudes and velocities permit us to separate express and anticipatory responses during the *Jumping Point* task. However, this is a very special task, and it is uncertain whether we can generalize the time windows for express saccades and the other proposed time ranges for fixation durations (see Figure 5) to other tasks. In Figure 12, the histograms of all fixation durations for our 6 standard tasks are shown.

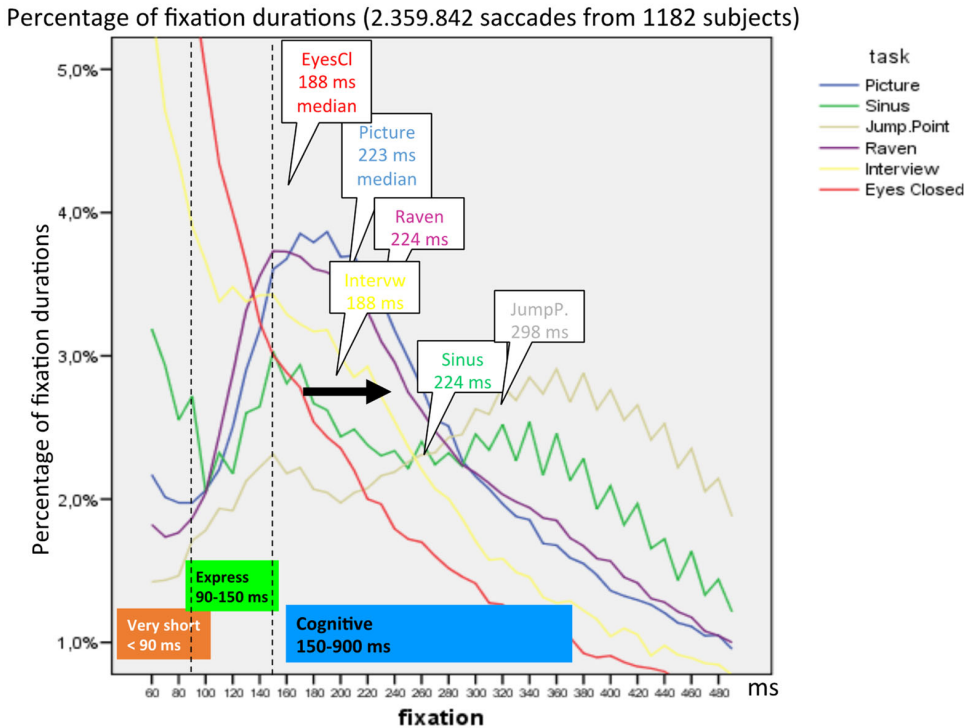


Figure 12 Fixation durations in our 6 standard tasks for all subjects. One can see three groups of curves: the *Raven* and the *Picture Inspection* task have their peaks at 130 to 220 ms, while the two tracking tasks, *Jumping Point* and *Sinus*, show double peaks from 300 to 400 ms and from 130 to 170 ms. The later peak at 300-400 ms is induced by the frequency of the target movement, which the subjects had to follow (see also figure 4). The third group of histogram curves stems from the interview tasks, *Interview* and *Eyes Closed*, during which subjects had no visual task to carry out. These curves show the highest values for short fixations, followed by an exponential decrease.

While the *Raven* and *Picture Inspection* tasks have their peaks in the *cognitive* and *express* range, the tracking tasks, *Sinus* and *Jumping Point*, have their peaks at the target frequency and additionally at approximately 150 ms. In contrast, the *Interview* and the *Eyes Closed* tasks in the no-visual-task group show no peaks in the *cognitive* and *express* range but the highest values for short fixation durations followed by an exponential decay. It seems noteworthy that some tasks have low values in the *very short* range (<90 ms) while the *Interview* and the *Eyes Closed* tasks do not. *Sinus*, *Picture Inspection*, *Raven* and *Jumping Point* – in this sequence – show decreasing values in the *very short* range. An obvious interpretation seems to be that more inhibition of gaze initiation is effective in these tasks and prevents irrelevant behavior. This is lacking in the tasks *Eyes Closed* and *Interview* and results in more early and short fixations.

Because amplitudes and velocities gave us hints as to the time window for express fixations in the *Jumping Point* task, we computed these parameters for our other standard tasks too (see Figures 13 and 15).

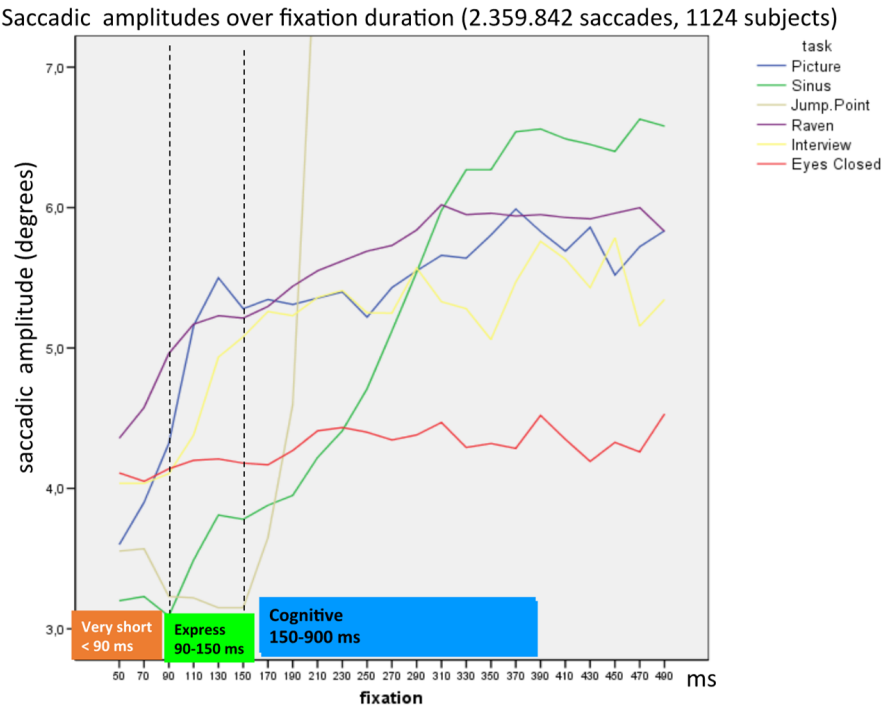


Figure 13 Amplitudes for all saccades (in degrees) over fixation duration (in ms) in our 6 standard tasks for all subjects. In all six tasks the median saccadic amplitudes are smaller in the *very short* range, increase in the *express* range (except in the *Jumping Point* task) and reach a plateau or increase further in the *cognitive* range. However, even in the *very short* range, amplitudes are larger than the microsaccade range (< 1 degree). The curve for *Jumping Point* shows a dramatic increase at the beginning of the *cognitive* range of fixations.

The amplitudes of the saccades were small in the *very short* range and increased in the *express* range, except in the *Jumping Point* task; here a dramatic increase began with the *cognitive* range at 150 ms. This seems to be a confirmation that at this point a new area begins. In the *express* range the amplitudes in this task were the lowest, which corresponds to the experience that small corrective saccades are often followed by undershooting response saccades. In the no visual task *Eyes Closed* the amplitudes remain nearly constant at 4 degrees. This may be a hint that the default condition for our inborn searching behavior that does not rest when we close our eyes is equipped with such an amplitude. This 3-fold range of our fovea may guarantee that new information in our nearby environment isn't overlooked but frees us from the strenuous looking for small details.

In Figure 14, a linear increase of fixation duration over saccadic amplitude in *Raven* and *Picture Inspection* is shown. When looking for the next object in the periphery, the visual system needs an additional 11 ms per one degree of distance for evaluation. Other research has reported a slightly different value: 8 ms /degree for the shift of attention over the visual field (Tsal, 1983). The slopes for the *Picture Inspection* and *Raven* increases were separated by some 40 ms. The fixation durations for *Picture Inspection* were greater, which seems to be caused by the lower contrast of the pictures in relation to the high-contrast line drawings in the *Raven*.

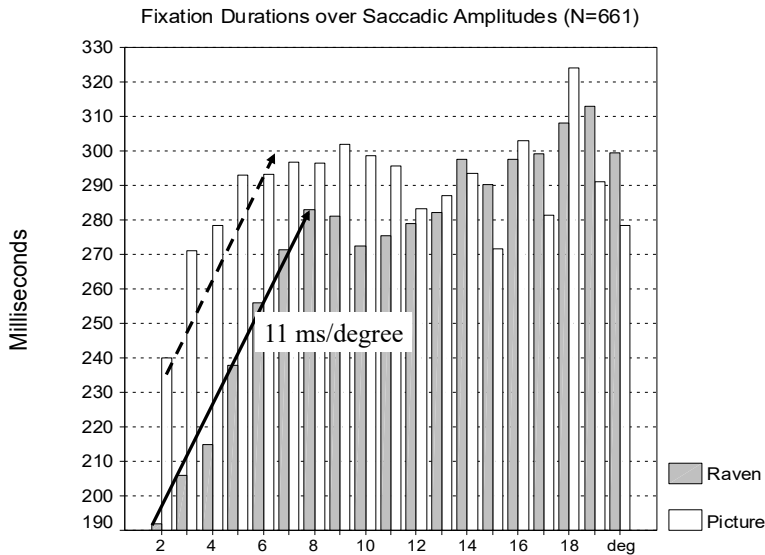


Figure 14 Fixation durations increase over saccadic amplitude with approximately 11 ms per degree distance until a plateau is reached, which occurs at 6 degrees in *Picture Inspection* and 8 degrees in *Raven*.

Looking at the (standardized) saccadic velocities (in degrees per second) over fixation duration (in ms) during our standard tasks (see Figure 15), one can see that the velocities were low in the *very short* range, increased in the *express* range (except for the *Eyes Closed* task) and showed different tendencies in the *cognitive* range.

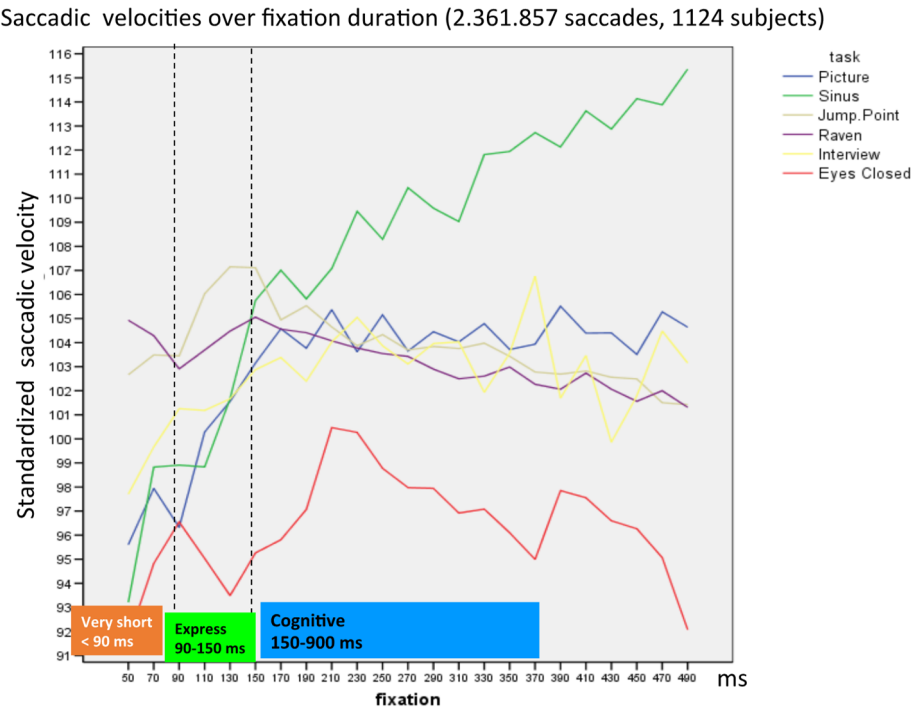


Figure 15 The (standardized) velocities for all saccades over fixation duration (in ms) in our 6 tasks were lowest in the *very short* range, increased in the *express* range (except for the *Eyes Closed* task) and showed different tendencies in the *cognitive* range.

The lowest velocity in the *Eyes Closed* task is partly caused by the braking effect of the closed lid on the eyeball. The tasks showed the lowest values in the *very short* range, increases in the *express* range and showed differences in the *cognitive* range. Here the *Sinus* task showed the highest values and further increases. This corresponds to the fact that subjects judged the continuous tracking of an accelerated moving point to be the most strenuous task, while *Jumping Point* was judged as less strenuous, although it had the same frequency. However, the task allowed rests between the jumps and did not need uninterrupted attention to the target. At exactly the beginning point of the *cognitive* range (150 ms) a plateau is reached by the velocity curves. This confirms the borderlines between the time windows.

Thus we have presented evidence that separation of fixation duration into three ranges - *very short*, *express*, and *cognitive* - is supported by different parameters of gaze behavior: the peaks in the distribution curves, the behavior of saccadic velocities, and saccadic amplitudes. We learned that the difficulty of the task as well as the IQ of the subjects had a small effect on fixation duration. As we can see in Figures 16 and 17, the age of the probands had a greater impact on fixation.

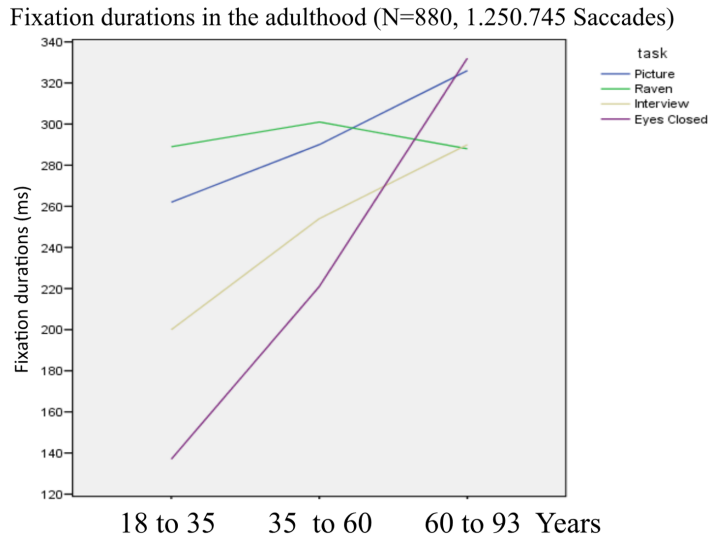


Figure 16 Median fixation duration in adulthood increases with the age of the subject. With the exception of the *Raven* task, fixation durations increased continuously over age.

Increasing age increased fixation duration in most tasks. The effect is dramatic: elderly subjects needed 60 ms more during *Picture Inspection* than middle-aged persons and 200 ms longer during *Eyes Closed*. It seems that mental processes take longer with age. That is exactly what the mental speed theory had described (Salthouse, 1996). Therefore, the reverse should also be found for youth. Becoming mature, i.e. an adult, speeds up mental processes, as seen in Figure 17.

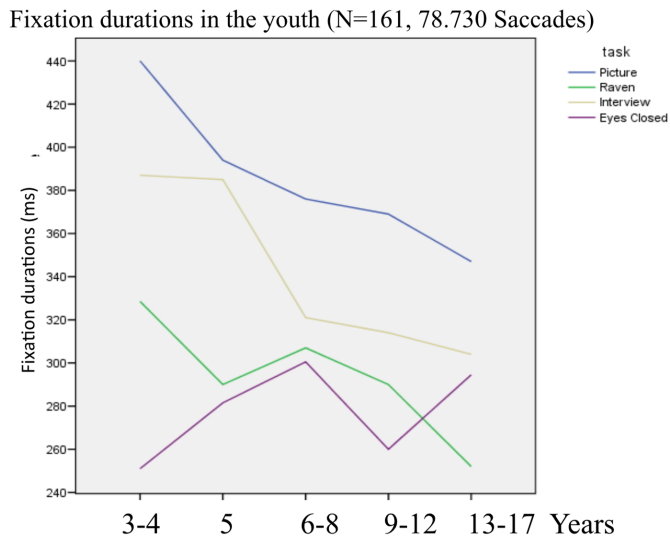


Figure 17 Median fixation duration in youth decreases with the age of the subject. The exception is the *Eyes Closed* task, in which the increasing speed of fixation durations did not occur.

Fixation duration in youth decreased when age increased: young children (3-4 years old) looked for 440 ms at a detail in the *Picture* before moving to the next one, while 14-year-old needed only 360 ms. Only the *Eyes Closed* task did not show such a systematic trend of increasing speed. However, in this task there is no picture for the visual system to process.

Until this point, we haven't examined the long and the *very long* fixations, our fourth category of fixation durations. All the fixation duration medians we have shown are in the *cognitive* range. But, this is a characteristic of medians: extremes are left out.

In a study of 129 drivers during a prolonged trip in a driving simulator we found that tired drivers showed more short fixations < 150 ms, less cognitive fixations (150 to 900 ms) and more very long fixations, i.e., longer than 900 ms (Schleicher, Galley, Briest & Galley, 2008). From these data we concluded that there must be a fourth category of very long fixations outside the *cognitive* window. Very long fixation durations are usually very rare, so there is not enough data for an undeniable functional interpretation. However, the non-cognitive connotation seems justified by the increase in very long fixations in the Falling Asleep Condition as well as their more frequent occurrence in very old people (our own unpublished data).

Discussion and Conclusion

We have attempted to provide an explanation for the apparently high variability of fixation durations by looking at the distribution curves of fixations during different tasks (Figures 2, 4, 5, and 12). This led us to believe that fixation is a process in time that has four ranges: (1) *very short fixations*, (2) *express fixations*, (3) *cognitive* and (4) *very long fixations*. We found evidence for the first three ranges in the histograms of the latencies, velocities, and amplitudes of the response saccades in the *Jumping Point* task (Figures 7, 10, 11). In Figures 13 and 15, the behavior of saccadic amplitudes and velocities from the other tasks over the fixation duration provided evidence that these ranges are valid for other tasks, too. Obviously, the task had an enormous influence on the distribution curves of fixation duration, much more than the value of the central tendency reveals (Figure 12). For example, the tasks *Raven* and *Sinus* show the same 224 ms median fixation duration but have very different distribution curves.

It has often been suggested that the phenomenon of fixation has to be understood as the interplay of two systems: the visual and oculomotor system on one side and the cognitive system on the other side. This is in accord with our results. We often found that lengthening the fixation duration was accompanied by signs of inhibition, for example fewer very short fixations (see Figure 12 and the related text), which signify attention and cognitive control. The task *Jumping Point* tolerates fewer irrelevant, very small gazes than *Sinus* and *Raven*, but during *Interview* small saccades do not disrupt the task and are disinhibited or allowed. Thus signs of inhibition can be absent in other tasks, and the oculomotor system can work in its default mode as a searching system with a high frequency of movements, for example in the *Eyes Closed* task. A closer look at Figure 12 shows that

the curve of *Interview* is shifted to the right for approximately 60 ms in relation to *Eyes Closed*. In other words, having an image (*Interview*) lengthened fixation compared with not having an image (*Eyes Closed*).

In a comprehensive study about the factors that guide the eyes while searching Pomplun (1998) reported a small difference in two nearly identical pictures. Pomplun suggested a model “which should generate fixations and saccades on the basis of assumed mental processes and their parameters (strategies, visual acuity, attention, working memory etc.” (p. 185). The inputs for the model generator were a) overall reaction time for finding the target, b) fixation durations, c) saccadic amplitudes, d) the probability of missing the target and some derived measures. He stated that basic research has found that fixation can be ambiguous in the sense that one gaze can eventually process two items if their distance is small enough that both items fall in the area covered per fixation. He reported that fixation duration increased with item intensity and in the last verification phase compared with the beginning exploration phase. Unfortunately, Pomplun used the mean of fixation duration as a measure of the central tendency. This is an inappropriate descriptor for the measurement of the central tendency; thus his reported values of most values over 200 ms seem to be overtaken by some very long fixation durations (see his Figure 3.3. on p. 56, where all fixations <100 ms have been excluded. This is another sign that most researcher are not interested in very short fixations). Interestingly, he proposed a category called ‘long fixations’, as we also did, but he used 500 ms (p. 56) as the critical value, while we used 900 ms as the critical value. Pomplun did not report if these values were excluded from the ‘mean’ computation. Thus Pomplun could not use the probabilities of different time ranges to draw conclusions about inhibition processes as we did.

What are the functional implications of dividing fixation duration into four ranges?

Firstly, very short fixations could not be controlled by cognitive, for example conscious mental processes. They are therefore often excluded from further analysis; but the proportion of fixations in this range can signify the inhibitory influence of the cognitive system. For example, it is minimal in *Eyes Closed* and maximal in *Jumping Point*.

Secondly, the *express* range seems to reflect simple cognitive processes, for example, “this is not the intended position – I must go further in this direction”. This is often the case in corrective saccades, but they can hardly be detected in most tasks with the exception of *Jumping Point*. Other examples of reduced costs of cognitive control are *express* range response saccades, which can be seen in decreased latencies when the jumps were expected in *Jumping Point*. However, these *express* reactions are not anticipatory saccades, which are characterized by smaller amplitudes and lower velocities, as we have seen. Often, *express* fixations seem to reflect simply the tendency of the vigilant oculomotor system to function as a high frequency search system and the absence of cognitive inhibition.

Thirdly, in the *cognitive* range, there is enough time for cognitive operations to detect matches or mismatches between the object seen and the expected object. If time is insufficient, the cognitive system allows a microsaccade, which does not remove the object from the fixation area. If the cognitive operation on the object is concluded, a new object outside the fixation area is grasped. However, this takes time, which we had calculated as 11 ms per degree (Figure 14).

Fourthly, although the consideration of fixation frequencies in the categories *express* range and *very short* range gives hints that the cognitive systems has inhibited the saccadic system, it contains no simple indications as to what the cognitive system is working on. For many years, researchers believed that scan path analysis could help, because it seems obvious that the sequence of different areas of fixation mirrors different mental content. However, the problem lies in the uncertainty as to whether a microsaccade was voluntarily controlled or happened automatically, for example to prevent the fading out of visual perception. Another ambiguity arises because longer fixation durations can suggest more difficult identification or detection processes or the incidence of two neighboring items being processed by one fixation. Thus the mental processes behind a sequence of fixations are complicated and cannot simply be deduced by oculomotor parameters (Schütz, Braun & Gegenfurtner, 2011).

References

- Aserinsky, E. & Kleitman, N. (1953). Regularly Occurring Periods of Eye Motility, and Concomitant Phenomena, during Sleep. *Science*, 118 (3062), 273-274. doi: 10.1126/science.118.3062.273
- Bronstein, A. M. & Kennard, C. (1987). Predictive eye saccades are different from visually triggered saccades. *Vision Research*, 27(4), 517-520. doi: 10.1016/0042-6989(87)90037-X
- Fischer, B. & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, 16(3), 553-567. doi: 10.1017/S0140525X00031575
- Galley, N. (1984). An activation paradigm using saccadic movements. In A. G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research*, (511-521). Amsterdam: Elsevier.
- Galley, N. (1987a). Saccadic velocity is regulated by the command mode of the subject and not by the stimulus parameters per se. In G. Lueer & U. Lass (Eds.), *Fourth European Conference on Eye Movements*, (127-129). Toronto: Hogrefe.
- Galley, N. (1987b). Saccadic velocity varies as a function of 'activation'. In J. K O'Regan & A. Lévy-Schoen (Eds.), *Eye movements - from physiology to cognition*, (154-155). Amsterdam: Elsevier.
- Galley, N. (1989). Saccadic eye movement velocity as an indicator of (de)activation. A review and some speculations. *Journal of Psychophysiology*, 3, 229-244.
- Galley, N. (1990). *Von Augen-blick zu Augen-blick. Eine experimentelle Psychologie der sakkadischen Augenbewegungen*. Habilitationsschrift. Philosophische Fakultät der Universität zu Köln, Deutschland.
- Galley, N. (1993). Augenbewegungen, Antizipation und Leistung. Auf dem Wege zu einem neuropsychologischen Modell der Konzentration. In J. Beckmann, H. Strang & E. Hahn (Eds.), *Aufmerksamkeit und Energetisierung. Facetten von Konzentration und Leistung*, (229-246). Göttingen: Hogrefe.

- Galley, N. (1998). An enquiry into the relationship between activation and performance using saccadic eye movement parameters. *Ergonomics*, 41(5), 698-720. doi: 10.1080/001401398186865
- Galley, N. (2001). Physiologische Grundlagen, Meßmethoden und Indikatorfunktion der okulomotorischen Aktivität. In F. Rösler (Ed.), *Enzyklopädie der Psychologie. Grundlagen und Methoden der Psychophysiologie. Biologische Psychologie*, (Vol. 4, 237-316). Göttingen: Hogrefe.
- Galley, N. (2010). Cognitive Development over the Life Span as Indicated by Eye Movements. In R. Zukauskienė (Ed.), *XIV European Conference on Developmental Psychology*, (33-41). Vilnius, Lithuania: Medimond.
- Galley, N. & Fritz, A. (1987). Augenbewegungen von Kindern mit unterschiedlich diagnostizierten Hirnfunktionsstörungen. *Zeitschrift für Kinder- und Jugendpsychiatrie*, 15, 342-354.
- Galley, N. & Galley, L. (1999). Saccadic latency and fixation durations as indicators of mental speed. In I. Mervielde, I. Deary, F. De Fruyt & F. Ostendorf (Eds.), *Personality Psychology in Europe*, (Vol. 7, 221-234). Tilburg: Tilburg University Press.
- Holmqvist, K., Nyström M., Andersson, R., Dewhurst, R., Jarodzka, H. & Van de Weijer, J. (Eds.) (2011). *Eye Tracking. A comprehensive guide to methods and measures*. Oxford, Oxford University Press.
- McPeck, R. M., Skavenski, A. A. & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision Research*, 40 (18), 2499-2516. doi:10.1016/S0042-6989(00)00102-4
- Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I. & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision*, 8(14), 21, 1-18. doi: 10.1167/8.14.21
- Pomplun, M. (1998). *Analysis and models of eye movements in comparative visual search*. Göttingen: Cuvillier.
- Radach, R., Heller, D. & Inhoff, A. (1999). Occurrence and function of very short fixation durations in reading. In W. Becker, H. Deubel & T. Mergner (Eds.), *Current Oculomotor Research: Physiological and Psychological Aspects*, (321-332). New York, N.Y.: Plenum Press. doi: 10.1007/978-1-4757-3054-8_46
- Raven, J. C., Court, J. & Raven, J. (1978). Coloured Progressive Matrices. CPM. *Raven-Matrizen-Test*. Weinheim: Beltz Test.
- Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Research*, 49, 2415-2441. doi: 10.1016/j.visres.2009.08.010
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403-428. doi: 10.1037/0033-295X.103.3.403
- Schleicher, R. & Galley, N. (2005). Saccadic velocity and duration as psychological variables I. Standardization procedure. *13th European Conference on Eye Movements*. Bern: University of Bern.
- Schleicher, R., Galley, N., Briest, S. & Galley, L. (2008). Looking tired? Blinks and saccades as indicators of fatigue in sleepiness warners. *Ergonomics*, 51(7), 982-1010. doi: 10.1080/00140130701817062

- Schütz, A. C., Braun, D. I. & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, 11(5), 9, 1-30. doi: 10.1167/11.5.9
- Schweigert, M. (2003). *Fahrerblickverhalten und Nebenaufgaben*. Dissertationsschrift. Fakultät für Maschinenwesen, Technische Universität München, Deutschland.
- Smit, A. C. & Van Gisbergen, J. A. M. (1989). A short-latency transition in saccadic dynamics during square-wave tracking and its significance for the differentiation of visually-guided and predictive saccades. *Experimental Brain Research*, 20, 64-74.
- Sommer, M. A. (1994). Express saccades elicited during visual scan in the monkey. *Vision Research*, 34 (15), 2023-2038. doi: 10.1016/0042-6989(94)90030-2
- Thier, P. & Ilg, U. J. (2005). The neural basis of smooth-pursuit eye movements. *Current Opinion in Neurobiology*, 15, 645-652. doi: 10.1016/j.conb.2005.10.013
- Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 9(4), 523-530.

5 Der Einfluss unterschiedlicher Gerechtigkeitsprinzipien auf Mehrheitsentscheidungen in einem spieltheoretischen Social-Good-Experiment¹

Claudia Biniossek¹ und Dirk Betz^{1,2}

¹ GESIS – Leibniz-Institut für Sozialwissenschaften

² ISS Institut für Soziologie und Sozialpsychologie

Zusammenfassung

Vor dem Hintergrund heterogener, teils widersprüchlicher experimenteller Befunde zum prosozialem Entscheidungsverhalten in großen Gruppen wird die Forschungsfrage gestellt, welchen Einfluss die finanzielle Schlechterstellung von unterschiedlich großen Subgruppen auf das Verhalten von Individuen hat. Hierzu werden Daten aus neun Varianten eines spieltheoretischen Social-Good-Großgruppenexperiments (Spiel 1: Hörsaalexperiment Uni Köln mit N=99, Laborexperiment im Kölner Laboratorium für Wirtschaftsforschung [CLER] mit N=162) mit 33 bzw. 27 Gruppenmitgliedern erhoben. Die Versuchspersonen müssen sich per Mehrheitsentscheid zwischen Gruppeninteresse und Eigeninteresse entscheiden. Die Messung behavioraler Komponenten wird kombiniert mit einer Erhebung der Entscheidungsgründe anhand von schriftlichen Survey-Fragen und qualitativen Interviews. Die Messung wird zudem um die Erhebung einer Spendenoption (Spiel 2; N=162) ergänzt. Es zeigen sich folgende Effekte: Akteure verhalten sich mehrheitlich eigennutzmaximierend. Jedoch wählen viele Individuen auch in großen Gruppen die distributiv faire Alternative im Rahmen der spezifischen Low-Cost-Hypothese. Nur wenn eine sehr kleine Minderheit stark benachteiligt wird, kommt es hochsignifikant zur Hilfe für die Benachteiligten im Sinne des Identifiable-Victim-Effekts. Das Modell der Frame-Selektion (MFS) erweist sich als aussagekräftig für die Erklärung, unter welchen Bedingungen Individuen ein bestimmtes der konkurrierenden Gerechtigkeitsprin-

1 Postprint: Biniossek, C., & Betz, D. (2018). *Der Einfluss unterschiedlicher Gerechtigkeitsprinzipien auf Mehrheitsentscheidungen in einem spieltheoretischen Social-Good-Experiment*. (GESIS Papers, 2018/20). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. doi: 10.21241/ssoar.58261

zipien auswählen. Jedoch kann die Annahme situationenüberspannender sozialer Präferenzen nicht bestätigt werden.

Schlüsselwörter: Rational Choice, Low-Cost-Hypothese, Modell der Frame-Selektion (MFS), sozialpsychologische Gerechtigkeitsforschung, experimentelle Spieltheorie

The Influence of Different Principles of Justice on Majority Voting in a Game Theoretic Social Good-Experiment

Abstract

Previous experimental findings about the decision-making behavior of individuals interacting in large groups tend to be contradictory. This essay focuses on the question of how (monetary) discrimination of various subgroups influences the individual decision-making process in large groups. For this purpose, data from nine variants of a game-theoretical social good group experiment with 33 respectively 27 group members were analyzed. Participants had to decide between group interest and self-interest by majority vote. The quantitative analysis of individual decision-making is complemented by a written survey and a qualitative interview. Furthermore, individuals have the opportunity to donate some of their post-game funds to a charity. The results are as follows: The majority of individuals act self-maximizing. However, a significant amount of individuals, even in large groups, choose the concept of equality as principle of justice within the framework of the specific low-cost hypothesis. Only if a very small minority is heavily disadvantaged, it is highly significant to help the disadvantaged group members in the sense of the identifiable victim effect. The model of frame selection (MFS) proves successful in explaining when individuals select one of several competing justice principles. However, it was not possible to confirm the assumption of social preference across situations.

Keywords: rational choice, low-cost hypothesis, model of frame selection (MFS), social justice research, experimental game theory

Einführung²

Der Einfluss unterschiedlicher Gerechtigkeitsprinzipien auf Mehrheitsentscheidungen und die Aggregation individuellen Entscheidungsverhaltens sind aufgrund ihrer hohen gesellschaftlichen Relevanz wie zum Beispiel im Hinblick auf populistische Volksbewegungen, Wahlen, oder die Bereitstellung von Kollektivgütern ein zentrales Thema in den Sozialwissenschaften. Empirische Befunde zum Einfluss der Gruppengröße auf das (Entscheidungs-)Verhalten zeichnen ein heterogenes Bild. Dem in vielen empirischen Untersuchungen nachgewiesenen Zusammenhang, dass es mit steigender Gruppengröße auch zu mehr eigennützigem Verhalten in Form von Trittbrettfahrerverhalten kommt, wie es von Olsons Theorie kollektiven Handelns (1965) prognostiziert wird und auch der standardökonomischen Prognose entspricht, stehen Befunde gegenüber, wonach die Gruppengröße nicht zwangsläufig den von der Public-Choice-Theorie prognostizierten Effekt hat (Isaac und Walker 1994). Saueremann (2017) konnte den großen Einfluss von distributiver Gerechtigkeit auf demokratische Entscheidungen anhand von Laborexperimenten aufzeigen. In einem Öffentliche-Güter-Spiel (ÖG-Spiel) von Weimann et al. (2014) sind beispielsweise sogar in großen Gruppen mit 60 bzw. 100 Gruppenmitgliedern die individuellen finanziellen Beiträge signifikant höher als in kleinen Gruppen mit acht Spielern³. Zudem sind bei Wiederholung des ÖG-Spiels über mehrere Runden in diesem Experiment die gleichen individuellen Verhaltensmuster unabhängig von der Gruppengröße identifizierbar. Im Einklang mit diesen neueren Ergebnissen gingen Wirtschaftswissenschaftler bis zum finanzwissenschaftlichen Paradigmenwechsel von der paretianischen Wohlfahrtsökonomik (Musgrave und Musgrave 1989) zu Rational Choice bzw. dem Public-Choice-Paradigma der Neuen Politischen Ökonomie implizit davon aus, dass die Gruppengröße keinen Einfluss auf das individuelle Entscheidungsverhalten entfaltet, da die ältere Wohlfahrtsökonomie den Einfluss der Gruppengröße auf die individuellen Kosten- und Nutzenkalküle nicht thematisierte. In der sozialwissenschaftlichen Forschung besteht eine Fülle empirischer Evidenz, die einen inversen Zusammenhang zwischen Gruppengröße und Verhalten zu belegen scheint: je größer die Gruppe, desto geringer die Tendenz zu individuellem Trittbrettfahrerverhalten. Dies zeigt sich, beispielsweise bei Spenden einer breiten Öffentlichkeit für Einzelschicksale (Fetchenhauer et al. 2010, S. 12-13), bei kollektiven Entscheidungen und Verhalten im Bereich Umweltschutz (Best 2009), bei der solidarisch finanzierten Krankenversicherung (Biniossek et al. 2007), Freiwilligen-Arbeit (Klößner 2016) und Hilfeleistung für Flüchtlinge (Karakayali & Kleist 2016). Auch lassen sich Kombinationen aus diesen drei Effektverläufen identifizieren, beispielsweise sinkende

2 Wir danken der Deutschen Forschungsgemeinschaft (DFG) und dem Lehrstuhl für Wirtschafts- und Sozialpsychologie am Institut für Soziologie und Sozialpsychologie (ISS) der Universität zu Köln für die finanzielle Förderung.

3 Bezeichnungen, die in der männlichen Form verwendet werden, schließen die weibliche Sprachform ein.

individuelle Beiträge nur bis zu einem Plateau wie bei Replikationen des klassischen Ringelmann-Experiments bei Kravitz und Martin (1986).

Eine Integration der empirischen Befunde zu Gruppengößeneffekten über Disziplinengrenzen hinweg erscheint nicht zuletzt aufgrund der sehr hohen Kosten pro unabhängiger Beobachtung von tatsächlich realisiertem Entscheidungsverhalten wünschenswert. Doch scheitert diese nicht nur an einer mangelnden wechselseitigen Sichtbarkeit der Untersuchungsergebnisse oder an unterschiedlichen methodischen Herangehensweisen, sondern auch an einem fehlenden einheitlichen theoretischen Rahmen, wie ihn seinerzeit Olsons Logik des kollektiven Handelns darstellte.

In der vorliegenden Studie werden die Prognosen unterschiedlicher theoretischer Ansätze aus Ökonomie, sozialpsychologischer Gerechtigkeitsforschung und Soziologie für das Entscheidungsverhalten von Individuen in großen Gruppen einander gegenübergestellt, um die Leistungsfähigkeit der Konzepte anhand eines spieltheoretischen Experiments zu vergleichen. Hierzu werden zu Varianten eines Social-Good-Großgruppenexperiments mittels unterschiedlicher Erhebungsinstrumente umfassend Daten generiert. Bezüglich der Übernahme empirischer Befunde aus der experimentellen Wirtschaftsforschung beschreitet der vorliegende Artikel einen dritten Weg neben Ablehnung (Streeck 2010) und einer unkontrollierten Übernahme ohne Absicherung der zugrundeliegenden mentalen Prozesse (Liebe et al. 2017)⁴.

Die Messung behavioraler Komponenten in Form von Verhaltensmustern im Stratiemodus (Brandts und Charness 2011; Mitzkewitz und Nagel 1993; Selten 1967) wird kombiniert mit einer Erhebung der Entscheidungsgründe anhand von Survey-Fragen und qualitativen Interviews und um die Erhebung einer individuellen Umverteilungsoption ergänzt. Grundidee des Spieldesigns ist die folgende spieltheoretische Entscheidungssituation (siehe Tabelle 1): Eine Gruppe mit mehreren Spielern muss sich per Mehrheitsentscheid zwischen zwei Alternativen entscheiden, die sich sowohl in der Höhe der Gruppengesamtauszahlung als auch im Grad der Ungleichheit der individuellen Geldauszahlung zwischen den einzelnen Personen unterscheiden (Spiel 1). In Alternative A wird eine relativ niedrige Gruppengesamtauszahlung absolut gleichmäßig (Bolton und Ockenfels 2003) bzw. relativ gleichmäßig (Biniossek und Fetchenhauer 2007) unter den Spielern verteilt, während in Alternative B eine höhere Gruppengesamtauszahlung relativ ungleichmäßig verteilt wird. Dieses Design bietet den Vorteil, dass für Varianten des Spiels bereits experimentelle Daten und Erklärungsansätze aus Sozialpsychologie und experimenteller Wirtschaftsforschung vorliegen. Im Gegensatz zu ÖG-Spielen ermöglicht das Voting eine Variation isolierter Einzelfaktoren, ohne dass sich die übrigen

4 Ohne empirische Erhebung der kognitiven Prozesse bleibt beispielsweise unklar, ob Chefärzte im experimentellen Diktatorspiel mehr Geld transferieren, weil sie prosozialer sind als statusniedrigere Krankenschwestern, wie von Liebe et al. (2017) behauptet, oder weil 10€ für Chefärzte nur Spielgeld darstellen und die experimentellen Befunde aus einem methodischen Artefakt durch Verstoß gegen die Induced Value Theorie resultieren (Smith 1976).

Modellparameter durch die Gruppengröße oder durch die Änderung der Anreizstruktur zu Trittbrettfahrerverhalten verschieben. Anhand einer nachgelagerten Individualentscheidung (n-Personen-Diktatorspiel) wird zusätzlich das Ausmaß, die Konsistenz und die Stabilität individueller sozialer Präferenzen überprüft (Spiel 2).

In allen empirisch erhobenen Varianten des hier analysierten Social-Good-Spiels (Bolton und Ockenfels 2003; Biniossek und Fetchenhauer 2007; Lotz und Fetchenhauer 2012; Steiniger et al. 2015) entscheidet sich die Mehrzahl der Spieler für die individuell auszahlungsmaximale Option. Obwohl die Mehrheit der Spieler die Alternative A mit relativ gleichmäßiger Gruppengesamtauszahlung präferiert (Biniossek und Fetchenhauer 2007), verhält sich somit lediglich eine (robuste) Minderheit konsistent zur geäußerten allgemeinen Einstellung. Bolton und Ockenfels (2003) erklären die Wahl außerhalb von Eigennutz mit einem trade-off zwischen Fairness und Effizienz. Sie können in ihrem „Game II“ im Strategiemodus zeigen, dass 31,9% der benachteiligten Personen („Arme“) die (gruppen-)effiziente ungleichverteilte Option wählen. 18,1% der am besten gestellten Personen („Reiche“) sowie 38,9% der finanziell zwischen „Armen“ und „Reichen“ liegenden Personen („Mittlere“) entscheiden sich für die gleichverteilte Alternative A, was bei Bolton und Ockenfels als Fairness bzw. Ungleichheitsaversion interpretiert wird⁵. Dies bedeutet, dass sich ein signifikanter Anteil der Spieler in allen drei Rollen gegen die individuell auszahlungsmaximierende Option entscheidet. Steiniger et al. (2015) weisen nach, dass die Wahl der gleicher verteilten Alternative signifikant mit höheren Werten auf der Ungerechtigkeitssensibilitäts-Skala korreliert, wobei benachteiligte Personen mit hoher Opfersensibilität, gemessen in Form von hohen Werten auf der standardisierten Skala für Opfersensibilität (JS_{victim} Werte), nicht aus dem Motiv Fairness, sondern aus purem Selbstinteresse die gleichverteilte Option wählen. Für die anderen Spieler, die sich nicht in der Opferrolle befinden, korreliert die Wahl der gleicher verteilten Alternative signifikant mit höheren Werten auf der allgemeinen Ungerechtigkeitssensibilitäts-Skala für „andere“ (JS_{others} Werte). Die Aversion gegen Ungleichheit scheint hier im Gegensatz zu den „Opfern“ ein wichtiges Motiv zu sein.

Dem in allen Studien robusten Ergebnis, dass Kollektiventscheidungen von Eigennutz und untergeordnet auch von Fairness bestimmt werden, stehen sehr heterogene Befunde bezogen auf individuelle Entscheidungsmuster, Motive und das Ausmaß an Fairness gegenüber. Bereits geringe Designänderungen wie z.B. Strategiemodus oder Änderung der Auszahlungsmatrix entfalten teilweise eine große Wirkung. Kritisch zu diskutieren ist die Übertragbarkeit der experimentellen Ergebnisse auf gesellschaftliche Zusammenhänge, wie sie von allen Autoren angenommen wird, zumal die behaupteten gesamtgesellschaftlichen Zusammenhänge

5 Eine Sekundärauswertung der Rohdaten von Bolton und Ockenfels (2003) ergibt für die individuellen Entscheidungsmuster im Strategiemodus (N=72) für 29,2% der Spieler das Muster BAA (Wahl der ungleicher verteilten Option B in der Rolle als „Reiche“, Wahl der gleicher verteilten Alternative A als „Arme“ sowie als „Mittlere“); 26,4% BBB; 25,0% BAB; 6,9% AAA; 6,9% AAB; 2,8% ABB; 1,4% ABA und 1,4% BBA.

nur in kleinen, nicht aber in großen Gruppen erhoben wurden und ein isolierter Großgruppeneffekt nicht ausgeschlossen werden kann.

Im Folgenden werden Prognosen aus Konzepten der experimentellen Wirtschaftsforschung, der sozialpsychologischen Gerechtigkeitsforschung und der soziologischen Handlungstheorie abgeleitet, um die Erklärungskraft der unterschiedlichen Ansätze bezüglich des Einflusses unterschiedlicher Faktoren wie Gruppengröße, Setting (Labor versus Hörsaal), negative Auszahlungen und Anteile von Spielern in den Subgruppen auf das Entscheidungsverhalten in großen Gruppen zu überprüfen (Spiel 1). Zudem wird die Konsistenz, bezogen auf Entscheidungsmuster, Entscheidungsgründe und redistributives Entscheidungsverhalten in einer nachgelagerten Entscheidungssituation gemessen (Spiel 2).

Ungleichheitsaversion in der experimentellen Wirtschaftsforschung

Neben Eigennutzmaximierung werden in der experimentellen Wirtschaftsforschung auch soziale Präferenzen als Verhaltensgründe modelliert. Dies umfasst Altruismus (Auszahlung an Mitspieler stiftet Nutzen) (Andreoni und Miller 2002) und Reziprozität (Wechselseitige Berücksichtigung der Auszahlung stiftet Nutzen) (Falk und Fischbacher 2006), Ungleichheitsaversion (Bolton und Ockenfels 2000; Fehr und Schmidt 1999)⁶, den experimentellen Vergleich der Bedeutung von Effizienz versus Ungleichheitsaversion für das Verhalten (Engelmann und Strobel 2004), aber auch komplexere Modelle, die simultan Effizienz und Reziprozität modellieren (Charness und Rabin 2002).⁷ Sämtliche Konzepte sind modelliert als Erweiterung von Wert-Erwartungs-Theorien um soziale Präferenzen (Nutzen), jedoch wird häufig davon ausgegangen, dass in spieltheoretischen Experimenten das Verhalten zumindest langfristig in Richtung reinem Egoismus konvergiert (Brosig et al. 2007).

Auf der Grundlage empirischer Beobachtungen, wonach soziale Präferenzen insbesondere in Verhandlungs- und Kooperationsspielen auftreten (Forsythe et al. 1994), jedoch nur das Niveau und nicht die Struktur der Einkommensverteilung unter den übrigen Mitspielern einen Einfluss auf Entscheidungen hat (Güth und Van Damme 1998) und von Versuchspersonen sowohl die eigene (finanzielle) Schlechterstellung als auch die Schlechterstellung anderer Personen abgelehnt wird, formulierten Bolton und Ockenfels (2000) das ERC-Modell, aufbauend auf Camerer und Thaler (1995) und strukturell ähnlich zu Fehr und Schmidt (1999, 2010).

6 In der Literatur haben sich aus diesen Grundkonzepten sozialer Präferenzen komplexe mathematische Modelle mit teilweise abstrakten Modellparametern (zum Beispiel Erev et al. 2009) entwickelt, die beispielsweise als Derivate der Cumulative Prospect Theorie (Tversky und Kahneman 1992) modelliert sind. Einen Überblick bieten Van Damme et al. (2014).

7 In Treshold-Öffentliche-Güter-Spielen konnten neben Trittbrettfahren auch Gründe nachgewiesen werden, die prosoziales Verhalten hemmen wie beispielsweise die Angst vor einem Verlust des eingesetzten Geldes, falls das Kollektivgut nicht erstellt wird (Granovetter 1978; Rapoport 1988; Rapoport und Eshed-Levy 1989).

Ungleichheitsaversion wird von Bolton und Ockenfels (2000) als Fairness-Term modelliert, der einen maximalen Wert annimmt, wenn ein Spieler den egalitären Anteil als arithmetischen Mittelwert aller Auszahlungen erhält. Somit ist die Verteilungsstruktur unter den anderen Spielern ohne Relevanz, was „zahlreiche experimentelle Beobachtungen erklären [kann]“ (Diekmann und Voss 2004, S. 26), insbesondere auch die Befunde zum Güth-van Damme-Spiel. Da die Gruppengröße im ERC-Modell weder auf den Nutzenterm noch auf den Fairnessterm einen isolierten Einfluss hat, sollen die Prognosen von Bolton und Ockenfels (2003) zum Entscheidungsverhalten und zu Entscheidungsgründen in ihrem Drei-Personen Social-Good-Spiel in vorliegender Untersuchung für große Gruppen übernommen werden. Aus der Ungleichheitsaversion in der Modellierung des ERC-Modells lassen sich die folgenden Hypothesen ableiten:

H1: Eine Voting-Alternative wird umso wahrscheinlicher gewählt, je geringer die komparative Ungleichheit ist (Konvergenz des Fairnessterms zum maximalen Wert) (H1a), je höher die Effizienz ist (Gruppengesamtauszahlung) (H1b), je höher der Eigennutz ist (eigene absolute Auszahlung) (H1c). In Anlehnung an die Axiome von Neumann und Morgenstern (1944) wird postuliert, dass die Verhaltensmuster auch zeitlich konsistent zu den Verhaltensmotiven sind.

Sozialpsychologische Gerechtigkeitsforschung

Die Formulierung der Hypothesen wird bewusst an den Anfang dieses Unterkapitels gestellt, da hierdurch die Grundproblematik der unterschiedlichen Granularität von Konzepten und notwendigen Brückenhypothesen zur Formulierung konkreter Verhaltensprognosen deutlich wird. Obwohl es sich bei Konzepten der distributiven Gerechtigkeitsforschung um Einstellungsmaße mit relativer Verhaltensferne handelt, lassen sich unter der Annahme des Korrespondenzprinzips (Ajzen und Sexton 1999) allgemeine Hypothesen für das Entscheidungsverhalten ableiten:

H2: Wenn in einer Voting-Spielsituation eine Ressource „vom Himmel fällt“, dann steht in einer Mehrheitsentscheidung das Equality-Prinzip bzw. distributive Gleichverteilung im Vordergrund.

Konkret sollte distributive Gleichheit als Einstellungsmaß im vorliegenden Untersuchungsdesign situativ im Vordergrund stehen und auch ein maßgeblicher Anteil der Spieler für die gleicher verteilte Option votieren (Verhalten). Hierbei bleibt jedoch unklar, ob und wie Fairnessentscheidungen bei einem Konflikt mit anderen Gerechtigkeitsnormen variieren. Diese Problematik soll im Folgenden verdeutlicht werden.

Die vor allem sozialpsychologische distributive Gerechtigkeitsforschung knüpft an (normative) philosophische bzw. juristische Gerechtigkeitsprinzipien an, die ethisches Verhalten und Wohlfahrt in einer Gemeinschaft adressieren. In der frühen sozialpsychologischen Gerechtigkeitsforschung standen normative Überlegungen mit einer Fokussierung auf Beitragsgerechtigkeit (Equity-Prinzip) im Vordergrund, dies führte zur Formulierung der Equity-Theorie (Adams 1965; Walster et al. 1973, 1976). Seit dem Aufsatz von Deutsch (1975) rücken vermehrt weitere Gerech-

tigkeitsprinzipien wie equality und need in den Vordergrund sowie die Regeln, in welcher Situation welches Prinzip gilt.

Schmitt und Montada (1982) zeigen als Erweiterung situationsspezifischer Gerechtigkeitskonzepte, dass individuelle Distributionsentscheidungen sowohl von der Situation als auch von der individuellen Präferenz für bestimmte Verteilungskriterien abhängig sind – ein Gerechtigkeitsprinzip „klebt“ somit nicht an situativen Randbedingungen, sondern wird vom Individuum ausgewählt. Dies steht im Widerspruch zu Konzepten aus der experimentellen Wirtschaftsforschung. Unter Einbeziehung weiterer Gerechtigkeitsprinzipien und Konzepte versucht Wiswede (2012) im Kontext der Konkurrenz unterschiedlicher Gerechtigkeitsnormen eine integrative Darstellung auf Grundlage der Lernpsychologie. Wiswede weist für zukünftige Forschung darauf hin, dass bei Überbezahlung eher kognitive Strategien und bei Unterbezahlung eher verhaltensaktive Strategien gewählt werden, die jedoch im Kontext einer Verflechtung unterschiedlicher (sozialer) Vergleichsniveaus zu analysieren sind.

Die bisher skizzierten Konzepte stellen Einstellungsmaße bzw. Persönlichkeitsmaße mit relativer Verhaltensferne dar. Infolgedessen wurde in der Konsistenzkontroverse (Schmitt 1990, 1993a, 1993b) die Frage nach der Handlungsrelevanz von Gerechtigkeitsprinzipien gestellt und im Konzept der Zentralität von Einstellungen die Einstellungsstärke als relevanter Faktor identifiziert. Jedoch mussten in der Folge auch diese Konzepte weiter differenziert werden, unter anderem in Form einer perspektivspezifischen Disposition des Erlebens von Ungerechtigkeit (Mikula 1993). Hierzu wurden Skalen zur Ungerechtigkeitssensibilität als Opfer, Täter, als Begünstigter oder unbeteiligter Beobachter entwickelt (Schmitt et al. 2009), die jedoch aufgrund ihrer Einstellungsnähe analog zu älteren Konzepten eine geringere Korrespondenz zum Verhalten aufweisen.

Jenseits der zentralen Konzepte der sozialpsychologischen Gerechtigkeitsforschung kommen zunehmend auch Biases und Heuristiken (Daumenregeln) zur Anwendung. Die Do-no-harm-Heuristik wird im Folgenden detaillierter dargestellt, weil sie einen Anwendungsbezug zum untersuchten Spieldesign aufweist und Probleme einer ex-ante Hypothesenbildung verdeutlicht. Die Do-no-harm-Heuristik ist als distributives Prinzip definiert „that it is wrong to harm some people in order to help others, even when the benefits outweigh the harm.“ (Baron 1995, S. 71). Die Wirkung der Do-no-harm-Heuristik konnte von Baron anhand einer quantitativen Umfrage mit 39 Studierenden auch für große Gruppen empirisch nachgewiesen werden. Eng damit verbunden ist der Identifiable-Victim-Effekt (Jenni und Loewenstein 1997; Schelling 1968) als Sonderfall der Do-no-harm-Heuristik, bei der Individuen versuchen, einen Schaden von kleinen Gruppen oder einzelnen Personen abzuwenden, deren Schicksal aus einer anonymen Masse heraussticht (Fetchenhauer et al. 2010). Hieraus folgt:

H3: Je stärker eine gut sichtbare kleine Minderheit benachteiligt ist, desto eher votieren die übrigen Kollektivmitglieder gegen diese Option.

Die weiter oben genannten Gerechtigkeitsprinzipien lassen sich konzeptuell nur teilweise mit der Do-no-harm-Heuristik verbinden, wenn zur Abwendung des Schadens für eine kleine Minderheit beispielsweise gegen die Prinzipien Gleichverteilung, Bedürfnisgerechtigkeit, Eigennutz und/oder die Wohlfahrt der Gesamtgruppe verstoßen wird. Ein Kernproblem in der aktuellen Gerechtigkeitsforschung ist der empirische Nachweis einer Vielzahl teils kontraintuitiver Phänomene, für die erst ex post Erklärungen und Mikrotheorien angeboten werden, weil die Prozesse zur Selektion angemessener Prinzipien und ihrer Handlungsrelevanz bisher nicht modelliert werden.⁸

Aktives Framing in der soziologischen Sozialtheorie

Ein wesentlicher Unterschied einiger soziologischer Konzepte gegenüber den oben dargestellten Ansätzen ist die explizite Modellierung der Logik der Situation sowie die Annahme aktiver Framing- und Wahrnehmungsprozesse. Situation und Persönlichkeitsdisposition definieren dagegen aus Perspektive passiver Wahrnehmung die Präferenzen und Entscheidungen von Individuen. In Analogie zu einem Computer wird das menschliche Gehirn in vielen dieser Konzepte als eine Art Rechenmaschine betrachtet, die persistent in gleichen Spielsituationen gemäß ihres Zeit überdauernden Persönlichkeitszuges (beispielsweise Ungleichheitsaversion [Bolton und Ockenfels 2000] oder Ungerechtigkeitssensibilität [Fetchnhauer und Huang 2004; Lotz et al. 2013]) konsistent zu identischen Entscheidungen kommt. Dies entspricht der Grundannahme des ökonomischen Verhaltensparadigmas und den Axiomen des Von Neumann-Morgenstern-Theorems (Von Neumann und Morgenstern 1944) mit der Annahme vollständiger, transitiver, stetiger und unabhängiger Präferenzen. Auch die Induced-Value-Theorie (Smith 1976) basiert auf der Annahme, dass mit der Auszahlungsmatrix bei angemessen hohen Geldbeträgen unabhängig von Raum und Zeit eine Nutzenstruktur induziert wird und sich Individuen in dieser induzierten Nutzenstruktur gemäß ihrer stabilen individuellen Präferenzen immer konsistent verhalten.

Die Annahme passiver Wahrnehmungsprozesse ist jedoch eine axiomatische Setzung, die aus Sicht einstufiger, passiver Entscheidungstheorien konkludent auch nicht empirisch erhoben werden muss. Demgegenüber analysieren Konzepte, die von aktiven Wahrnehmungsprozessen (Schwarz 1985; Stocké 2002), paralleler neuronaler Reizverarbeitung (Divergenz) und synthetischer statt analoger Integration von Bewusstseinsinhalten (Kandel et al. 2012) ausgehen, die Bedingungen, unter denen bestimmte Schlüsselreize bei einem Individuum in den Vordergrund treten

8 So konnte beispielsweise in einem Gedankenexperiment zur Verteilung von Grapefruits anhand des Vitamingehalts ein unerwartetes Umschlagen von distributiver Gleichverteilung der Vitaminwirkung zum Effizienzprinzip nachgewiesen werden (Bar-Hillel und Yaari 1993). Jedoch werden in dieser Studie nicht die Mechanismen spezifiziert, unter welchen Randbedingungen es zur Selektion eines alternativen Gerechtigkeitsprinzips kommt.

und spezifische Wahrnehmungs- und Verhaltensschemata aktivieren (Esser 1996). Sozialwissenschaftliche Entscheidungstheorien wie das Modell der Frame Selektion (MFS) (Esser 1996, 2001, 2010; Esser und Kroneberg 2015; Kroneberg 2005, 2011, 2014), aber auch das Modell von Lindenberg und Steg (2013) gehen von zweistufigen Entscheidungsprozessen aus. Hierbei weist das MFS strukturelle Ähnlichkeit zur Cumulative-Prospect-Theorie auf, geht aber im Gegensatz zu Kahnemans passivem Framing in der Editing-Phase (Stocké 2002) davon aus, dass sich Individuen in einer vorgelagerten aktiven Framing-Phase zunächst zwischen verschiedenen gedanklichen Modellen entscheiden – beispielsweise zwischen einem Idealmodell (perfekte Ehe, Religionsgemeinschaft, Beiträge zu öffentlichen Gütern in der ersten Runde von ÖG-Spielen oder Solidargemeinschaft der gesetzlichen Krankenversicherung) und einem alternativen Modell, das beispielsweise durch eine Krise ausgelöst wird (Reframing durch Ehekrise, Enttäuschung über Religionsführer [Festinger et al. 1956], Feedback ab der zweiten Runde im ÖG-Spiel über Trittbrettfahrer bzw. Berichte über Sozialbetrug durch GKV-Mitglieder), in dem nach Kosten und Nutzenaspekten bewertet wird (zweckrationales Verhalten)⁹.

Ein Kernproblem für ein entsprechendes Untersuchungsdesign ist der Mangel an geeigneten (z. B. qualitativen) Instrumenten zur Messung von Wahrnehmungsinhalten und zur Bestimmung gedanklicher Modelle sowie deren Triggerreizen. Bisher übliche Erhebungsinstrumente wie die Einstellungsmessung weisen eine relativ geringe Korrelation zum Verhalten auf (Six und Eckes 1996). Empirische Anwendungen von Framing-Konzepten referieren auf „starke“ gesellschaftliche Modelle, beispielsweise in Form von Normen (Best 2009; Lindenberg und Steg 2013) oder Ehe (wertrationales versus zweckrationales Verhalten) (Esser 1996), die eine beste Approximation an „tatsächliche“ gedankliche Modelle darstellen. Tutic (2015) schlägt vor, das Messproblem durch die Axiomatisierung von Verhaltenstheorien zu überwinden. In der vorliegenden Untersuchung wird demgegenüber der Versuch unternommen, das Problem anhand einer geeigneten Operationalisierung zu lösen und die Logik der Situation anhand der Messung von Verhaltensmustern (Strategiemodus) mit einer anschließenden Befragung der Spieler zu ihren Entscheidungsgründen zu erheben.

Gerechtigkeit im Modell der Frame-Selektion (MFS)

Aus Perspektive des MFS kann es über unterschiedliche Prozesse zu uneigennützigem Abstimmungsverhalten kommen: Reframing aufgrund einer unter Fairnessaspekten wahrgenommenen Situation, heuristisch-automatische Wahl der uneigennützigen Option (beispielsweise Mülltrennung bei Best, 2009), faires Verhalten aus Mangel alternativer Handlungsskripte (z. B. Trinkgeld im Restaurant trotz miserab-

⁹ Ansätze, die eine Logik der Situation modellieren, sind unmittelbar anknüpfungsfähig an Reinhard Seltens Konzept aufquellender Bewusstseinsebenen bzw. „3 level theory of decision making“ (1978) und die von Vernon Smith (2010) aufgeworfene Frage, wie Spieler eine Experimentalsituation perzipieren.

lem Service) sowie einer rationalen Selektion der fairen Votingoption aufgrund ihres (sozialen) Nutzens (analog zur einfachen Low-Cost-Hypothese [LCH]). Das MFS lässt prinzipiell sämtliche Wege (sowie Kombinationen dieser Prozesse) zu und modelliert diese explizit. Wir beschränken uns in vorliegender Arbeit auf Reframing und seine Kernkomponente Match.

Ausgangsgedanke des Modells der Frame-Selektion (MFS) ist eine Situation, in der ein spezifisches gedankliches Modell vorherrscht. Die Labortreatments des vorliegenden Experiments wurden am Kölner Laboratorium für Wirtschaftsforschung (CLER) durchgeführt, das die Versuchspersonen mit der Selbstdarstellung „Spielend Geld verdienen“ rekrutiert. Das grundsätzliche Motiv, aus Eigennutz Geld zu verdienen, wird somit durch die Eigenwerbung des Labors verstärkt, so dass hier zum Ausgangszeitpunkt t_0 vom gedanklichen Modell „Geldverdienen im Labor“ ausgegangen werden soll. Die axiomatisch gesetzte Annahme, dass alle Spieler diesem Frame folgen, wird in vorliegender Arbeit empirisch nicht überprüft. Zwar stellt diese Annahme eine starke Vereinfachung dar, da viele Teilnehmer auch durch Hinweise in Lehrveranstaltungen oder Mundpropaganda auf das Labor aufmerksam werden, andererseits wird durch die im Vorfeld angekündigte angemessene Entlohnung entsprechend der Induced Value Theorie (Smith, 1976) bestärkt, dass hier „Geldverdienen im Labor“ im Vordergrund steht. Das gedankliche Modell „Geldverdienen im Labor“ kann zusätzlich noch dadurch verstärkt werden, dass Spieler im Labor in einer anonymen Umgebung entscheiden als im Hörsaal, und sich daher dort eigennütziger verhalten. In diesem Frame stehen eigennützige Motive im Vordergrund und Entscheidungen werden zweckrational getroffen. Das gedankliche Modell, durch zweckrationale Entscheidungen spielend Geld zu verdienen, wird daher nachfolgend definiert als Frame „Zweckrational Geld verdienen“ und mit „ZG“ abgekürzt. Da komplexe Prozesse mit Modus- und Skriptselektion nicht im Zentrum der Analyse stehen und die Leistungsfähigkeit der zentralen Parameter des MFS aufgezeigt werden soll, folgen die Modellgleichungen analog der formalen Ableitung zur perfekten Ehe [Esser 2001, 2002a, 2002b, 2003]). Formal gilt für den Erwartungsnutzen des Modells „Zweckrational Geld verdienen“: F_{ZG} . Dieser Frame soll hier als nahezu „unverbrüchlich“ festgelegt werden. Dies bedeutet in der Logik des MFS, dass für den Grad des Matches ≈ 1 gilt.

Die Selektion eines Modells (Phase 1) definiert den Rahmen für Verhaltensentscheidungen (Phase 2) und besteht aus zwei Komponenten: Der Passung m des Frames zu einer Situation und dem subjektiv erwarteten Nutzen U der Wahl dieses Frames. Grundsätzlicher Unterschied zur Cumulative-Prospect-Theorie (Tversky und Kahneman 1992) und ihren Varianten ist somit eine aktive Frameselektion nach den Prinzipien der WE-Theorie (Stocké 2002). Hierbei wird die reizgebundene („bottom up“) Komponente „Abwesenheit von Störungen u “ eines Frames multiplikativ mit der Komponente „Zugänglichkeit a “ und der konzeptgebundenen („top down“-)Komponente „Existenz e “ des Frames verknüpft. Der Match bzw. der Grad m der Geltung eines Modells ist hierbei nach Esser formal definiert als: $m = a \cdot e \cdot u$. Der Match m steigt beispielsweise, wenn der Spieler in der Vergangenheit an Laborexperimenten teilgenommen hat (Zugänglichkeit a des Frames ZG), der aus dem

Laborexperiment erwartete Verdienst bereits vor der Teilnahme konkret verplant wurde (Existenz e der mit „Zweckrational Geld verdienen“ assoziierten Objekte) und eine Abwesenheit von Störungen u, beispielsweise das Empfinden von Ungerechtigkeit. Zum Zeitpunkt t_1 sind die Spieler durch die konkrete Spielsituation und die Auszahlungsmatrix mit neuen Informationen konfrontiert, die zu einer neuen Interpretation der Situation führen können. Zunächst ist zu analysieren, ob und wie sich die Ausgangssituation t_0 spielend „Zweckrational Geld verdienen“ durch das Treatment im Zeitpunkt t_1 für Individuen verändert.

Ein Wechsel der subjektiven Definition einer Situation wird im MFS als Reframing bezeichnet. Das MFS formuliert die Bedingungen für das Reframing: Zuerst muss ein mögliches alternatives gedankliches Modell zum Ausgangsmodell existieren. Zur Vereinfachung definieren wir die Vielzahl an Schemata, die dem gedanklichen Modell „Zweckrational Geld verdienen“ zugeordnet werden können, als „Frame-Universum Zweckrationalität“ mit der konkreten Ausprägung des Frames „Zweckrational Geld verdienen“ F_{ZG} . Das alternative gedankliche Komplementärmodell wird in Anlehnung an Max Weber definiert als „Frame-Universum Wertrationalität“ mit dessen Ausprägung als *wertrationaler Frame* F_{WR} . Es soll gelten:

$$EU(WR) = (1 - m)U_{WR}.$$

Analog zu Wert-Erwartungs-Theorien ergibt sich ein Wechsel des gedanklichen Modells (Reframing), wenn das EU-Gewicht des alternativen Frames $_{WR}$ (wertrational Entscheiden) größer ist als das Gewicht des Ausgangsframes $_{ZG}$ (zweckrational Geld verdienen). Als Bedingung für ein Reframing ergibt sich somit: $EU(WR) = (1 - m)U_{WR}$

Hieraus folgt $(1 - m)U_{WR} > mU_{ZG}$ und $\frac{U_{WR}}{U_{ZG}} > \frac{m}{(1 - m)}$.

Der „Grad der Unempfindlichkeit der Modell-Selektion gegen Variationen in den Parametern“ wird hierbei als Salienz definiert (Esser 2001). Formal ergibt sich:

$$S_{ZG} > \left(\frac{m}{(1 - m)} \right) - \left(\frac{U_{WR}}{U_{ZG}} \right).$$

Je unterschiedlicher die EU-Gewichte beider Frames sind, desto kleiner ist die Wahrscheinlichkeit, dass die Reframing-Schwelle erreicht wird. Bei einer hohen Salienz von Frame ZG bleibt somit das Modell „Zweckrational Geld verdienen“ im Vordergrund. Hieraus ergibt sich:

H4: Individuen entscheiden im Labor zu einem höheren Grad eigennützig als im Hörsaal.

Niedrigkostensituation im Rahmen des MFS

Im Rahmen von Arbeiten zur Low-Cost-Hypothese (Best und Kroneberg 2012) im Anwendungsbereich Kollektivgüter (Quandt und Ohr 2004) und Umweltverhalten¹⁰ (Best 2009; Diekmann und Preisendörfer 1992; Diekmann und Preisendörfer 1998; Diekmann und Voss 2004) konnte gezeigt werden, dass Entscheidungsverhalten in großen Gruppen in Niedrigkostensituationen zu einem höheren Grad durch Gerechtigkeitsmotive geprägt ist als in High-Cost-Situationen, in denen der Anwendungsbereich von „harten“ Modellen liegt (Diekmann und Voss 2004).

Zur Entstehung fairen Verhaltens argumentiert die LCH im Rahmen eines weiten Rational-Choice-Paradigmas, dass der Zusammenhang zwischen Einstellungsstärke und Verhalten durch die Kosten des Verhaltens moderiert wird und dass der subjektiv erwartete Nutzen moralischen Verhaltens in Low-Cost-Situationen die relativ geringen Kosten übersteigen kann.

Die Gerechtigkeitsüberzeugung fungiert also als „Zünglein an der Waage“ [Diekmann 1996, S. 107]). Da ein additives Einwirken von Kosten- und Nutzenaspekten auf faires Handeln das Umweltverhalten nicht vollständig erklären kann, wurde dieser theoretisch modellierte einfache Effekt von vielen Autoren um einen weiteren, gleichgerichteten Interaktionseffekt ergänzt. Diese spezifischen Versionen der LCH (Best und Kroneberg 2012) gehen davon aus, dass sich der Gesamteffekt in Niedrigkostensituationen aus zwei gleichgerichteten Effekten zusammensetzt: Dem unmittelbaren Nutzen aus der Befriedigung moralischer Überzeugungen und einem Interaktionseffekt zwischen Niedrigkostensituation und der Entscheidung. Die meisten Autoren argumentieren über eine situationsinduzierte Nutzenänderung, d.h. die Verschiebung der Nutzenkurve [Diekmann und Preisendörfer 1998], eine marginalistische Betrachtung bezüglich des abnehmenden Grenznutzens sozialer Anerkennung für „instrumentell“ moralisches Verhalten [Braun und Franzen 1995]) oder über eine Ausblendung von Kosten-Nutzen-Kalkülen in Low-Cost-Situationen (modelliert im Rahmen des MFS als periphere Route bzw. as-Modus [Best und Kroneberg 2012]). Die Argumentation von Kühnel und Bamberg (1998a, 1998b) über eine kosteninduziert unterschiedliche Wahrnehmung der Logik der Situation lässt sich ebenfalls den spezifischen LCH zuordnen und kann als Vorläufer einer Modellierung aktiver Framingprozesse interpretiert werden.

Im Folgenden wird die Wirkung von Niedrigkostensituationen auf die Modellparameter des MFS und das Reframing modelliert. Während oben bereits der Gesamtframe dargestellt wurde, entfaltet die Low-Cost-Situation insbesondere auf den Sub-Frame der finanziell zwischen „Armen“ und „Reichen“ liegenden Personen („Mittlere“) ihren Einfluss, da hier die Wahl eines wertrationalen Frames relativ wenig kostet. Für den Frame F_{WR} besteht eine besonders leichte Zugänglichkeit a sowie die Abwesenheit von Störungen u , beispielsweise in Form des Verlusts einer größeren Geldsumme. Das Streben nach konsistenten Antworten und Verhalten über

10 Jedoch konnte in vielen Bereichen wie zum Beispiel zum solidarischen Krankenversicherungssystem (Biniossek et al. 2007) Fairness auch in High-Cost-Situationen nachgewiesen werden.

alle drei Personengruppen wirkt simultan jedoch in umgekehrter Richtung auf eine Anwesenheit von Störungen u bezüglich F_{WR} , jedoch sollten als Gesamteffekt insgesamt weniger Störungen bestehen. Gleichzeitig steigt der relative Nutzen für ein Reframing, weil wertrationales Verhalten das Verschenken von Geld leichter macht, Regret-Effekte verhindert und hier relativ kostengünstig ist. Auf den Frame F_{ZG} wirkt die Low-Cost-Situation in umgekehrter Richtung, jedoch in stark abgeschwächter Form, da sich die Existenz e und Zugänglichkeit a kaum ändern.

In diesem geschwächten Gesamtrahmen bieten Niedrigkostensituationen die Gelegenheit, für wenig Geld den „warm glow of giving“ (Andreoni, 1989) zu erleben. In einer Niedrigkostensituation können hierbei leicht Schwellen erreicht werden, diese sozialen Bedürfnisse zu realisieren. Hier lässt sich als Hypothese formulieren:

H5: Je geringer die Kosten für die Realisierung einer als gerecht empfundenen Alternative sind, desto stärker sind die Auswirkungen auf den Match und den Nutzen von F_{WR} und desto höher ist die Wahrscheinlichkeit eines Reframings und somit der Wahl der als gerecht empfundenen Alternative.

Intersituative Konsistenz des Verhaltens und Kritik an der Methode der experimentellen Spieltheorie

Einige Autoren in den Sozialwissenschaften stehen der experimentellen spieltheoretischen Methode kritisch gegenüber (Beckert und Streeck 2008; Streeck 2010). Diese Kritik richtet sich nicht auf die oben erwähnte Axiomatisierungs- versus Operationalisierungsproblematik, sondern auf die mangelnde externe Validität und zweifelhafte Übertragbarkeit der Ergebnisse auf reale soziale Phänomene. Diese Argumente werden von spieltheoretischen Sozialwissenschaftlern zurückgewiesen (Kittel 2015). In spieltheoretischen Laborexperimenten entscheiden Individuen über Geldsummen, die mindestens dem üblichen Stundenlohn der Versuchsperson entspricht. Es handelt im Labor um echtes Entscheidungsverhalten, das sich selbst in High-Cost-Situationen als robust erweist. Um eine breitere Datengrundlage zur Beurteilung der Güte experimenteller Daten zu schaffen und um die temporale sowie inhaltliche Konsistenz der in Spiel 1 generierten Entscheidungsmuster und Verhaltensgründe zu explorieren, wird in den Laborbedingungen nach Abschluss von Spiel 1 ohne Vorankündigung Spiel 2 durchgeführt. In diesem erhalten Spieler die Möglichkeit, anhand eines N-Personen-Diktatorspiels (1) das Maß an Gleichheit innerhalb ihrer Gruppe nachträglich zu korrigieren, falls die Mehrheitsentscheidung nicht ihren Verteilungspräferenzen entspricht, und/oder zusätzlich (2) eine externe Umverteilungsoption zugunsten der Hilfsorganisation „Ärzte ohne Grenzen“ zu wählen. Anschließend werden die Spieler nach dem Grund ihrer Entscheidung gefragt um empirisch abzusichern, ob die Individuen Spiel 2 auch tatsächlich als Möglichkeit zur Realisierung prosozialer Motive identifizieren. Bezüglich der temporalen sowie inhaltlichen Konsistenz könnte argumentiert werden, dass in Spiel 1 Effizienz aus spieltheoretischer Sicht eine Rolle spielt, in Spiel 2 aber nicht. Die durch das Experimentaldesign intendierte Überprüfung intertemporaler Konsistenz adressiert jedoch auf die

Personen, die zuvor in Spiel 1 dreimal konsistent die faire Option gewählt haben. Für diese Spieler sollte Effizienz und psychologische Einkommenseffekte wie Mental Accounting (Thaler 1999) nicht relevant sein. Auch sozialpsychologische Moral-Licensing-Effekte (Sachdeva et al. 2009), dass Menschen weniger abgeben, wenn sie selbstwertförderlich bereits zuvor etwas Gutes getan haben, sollten nicht auftreten, da nur Gruppen mit der Entscheidung für die ungleiche Option analysiert werden, in denen das faire Verhalten noch nicht realisiert wurde. Zudem konnte der Moral-Licensing-Effekt in Metaanalysen nicht bestätigt werden (Blanken et al. 2014; Blanken et al. 2015). Unter der Annahme intertemporaler und inhaltlicher Konsistenz lässt sich zusammenfassend folgende Hypothese ableiten:

H6: Wenn es in Spiel 1 zu einem präferenzkonträren Mehrheitsentscheid kommt, dann realisieren Individuen ihre Verteilungspräferenzen aus Spiel 1 im anschließenden Spiel 2.

Methode

Versuchsdesign

Die oben dargestellten Hypothesen sollen anhand von Varianten des Untersuchungsdesigns von Bolton und Ockenfels (2003) überprüft werden, das zudem von 3 Gruppenmitgliedern auf 27 Gruppenmitglieder¹¹ (Laborexperimente) bzw. 33 Gruppenmitglieder (Hörsaalexperimente) erweitert wurde. Im Folgenden ist das Basisdesign des hier durchgeführten 27-Personen-Laborexperiments (Strategiemodus) dargestellt. Alle 27 Spieler einer Session sind Mitglied derselben Gruppe. Von den 27 Personen in dieser Gesamtgruppe sind 9 Gruppenmitglieder in der Subgruppe der „Reichen“, die nachfolgend als P1 bezeichnet werden. 9 Mitglieder sind in der Subgruppe der „Armen“ (P2), die restlichen 9 Mitglieder sind in der Subgruppe der „Mittleren“, nachfolgend als P3 bezeichnet. Die logische Verdrehung der Reihenfolge der Kodierung P1 = Reich, P3 = Mitte, P2 = Arm erfolgt analog Bolton und Ockenfels (2003) sowie Biniossek und Fetchenhauer (2008), um eine bessere Vergleichbarkeit herzustellen. In Tabelle 1 sind entsprechend den Spielinstruktionen die individuellen Auszahlungen je Versuchsperson (Vpn) im Basisdesign Laborexperiment abgebildet.

11 Die maximale Kapazität von Labor und Hörsaal sowie das Untersuchungsdesign definieren die Anzahl der Gruppenmitglieder.

Tabelle 1 Auszahlungsplan Basisdesign Laborexperiment (Reihenfolge analog Bolton und Ockenfels 2003)

	Alternative A	Alternative B
P1 („Reiche“)	7 € (sieben Euro)	13 € (dreizehn Euro)
P2 („Arme“)	5 € (fünf Euro)	1 € (ein Euro)
P3 („Mittlere“)	6 € (sechs Euro)	7 € (sieben Euro)

Beispiel: Wird eine Vpn als Mitglied der Subgruppe P1 („Reiche“) ausgelost, und entscheidet sich die Gesamtgruppe aller 27 Versuchspersonen mehrheitlich für Alternative A, dann erhält jede Vpn der Subgruppe P1 insgesamt 7 Euro. Dies bedeutet, dass an die 9 Mitglieder der Subgruppe P1 insgesamt 63 Euro ausgezahlt werden. Die Gesamtgruppe aller 27 Versuchspersonen erhält insgesamt 162 Euro.

Jede Vpn entscheidet zunächst im Strategiemodus (level playing field mode) für jede der drei potentiellen Personenrollen, ob sie die Option A (relative Gleichverteilung) oder B (höhere Gruppengesamtauszahlung bei ungleicher Verteilung) wählt. Erst anschließend wird jedem Spieler seine ihm per Los zugewiesene Personenrolle mitgeteilt.

Die Auszahlung jedes einzelnen Spielers hängt davon ab, welche der beiden möglichen Auszahlungsalternativen (A oder B) die Gesamtgruppe wählt und in welcher Subgruppe er ist. Welche Auszahlungsalternative die Gesamtgruppe wählt, ergibt sich per Mehrheitsentscheid.

Insgesamt werden die 9 Designvarianten jeweils einmal gespielt (Spiel 1). Nach Entscheidung der Spieler im Strategiemodus wird anhand von qualitativen Interviews der wichtigste Entscheidungsgrund in jeder Personenrolle erhoben. In den Labortreatments werden zusätzlich in schriftlichen Survey-Fragen bezogen auf die Gesamtentscheidungssituation beispielsweise die Dimensionen Eigennutz, Fairness und Effizienz erfasst. Im Anschluss an Spiel 1 werden in den Laborexperimenten das Ausmaß, die Konsistenz und die Stabilität individueller sozialer Präferenzen anhand eines n-Personen-Diktatorspiels (Spiel 2) überprüft, indem jedem Spieler die Möglichkeit gegeben wird, sein Einkommen entweder zu behalten, oder beliebige Teilsummen seines Einkommens an die Hilfsorganisation Ärzte ohne Grenzen, an die Subgruppe der „Armen“, an die Subgruppe der „Mittleren“ und/oder an die Subgruppe der „Reichen“ umzuverteilen. Hierbei wird den Probanden erst nach Abschluss von Spiel 1 mitgeteilt, dass im Anschluss Spiel 2 durchgeführt wird, indem die Instruktionen für die Umverteilungsoption auf dem Computermonitor eingeblendet werden.

Das Basisdesign wird in insgesamt 9 Varianten gespielt. Zur Überprüfung der Hypothesen wird die Erhebungsmethode (sechs Laborexperimente mit je 27 Spielern versus drei Hörsaalexperimente mit je 33 Spielern), die Auszahlungsmatrix (in drei der sechs Laborexperimente erhält P2 in Alternative B die negative Auszahlung

minus 1 Euro) sowie die Anteile der Personen P1, P2 und P3 variiert. In Tabelle 2 sind für die neun Treatments Personenanteile und Kennzahlen zu Effizienz- und Gerechtigkeitsmaßen (aufsteigend geordnet nach Effizienzquotient) dargestellt.

Tabelle 2 Variation von Gleichverteilung und Effizienz über die Treatments

Treatment	Gruppen zusammensetzung Anzahl Personen P1, P2 und P3	Effizienzmaße				Fairness-Maße	
		Gruppengesamt- auszahlung in Euro (Cake)				Gini- Gerechtigkeits- koeffizient*	
		Option A in €	Option B in €	Effizienz- quotient A/B	Effiziente Option	Option A	Option B
HS_3Arme	15*P1; 3*P2; 15*P3	210	303	0,69	B	0,052	0,216
3Arme	12*P1; 3*P2; 12*P3	171	243	0.70	B	0,055	0,230
3ArmeMinus1€	12*P1; 3*P2; 12*P3	171	237	0.72	B	0,055	0,259
HS_Gleich	11*P1; 11*P2; 11*P3	198	231	0,86	B	0,074	0,381
Gleich	9*P1; 9*P2; 9*P3	162	189	0,86	B	0,074	0,381
GleichMinus1€	9*P1; 9*P2; 9*P3	162	171	0,95	B	0,074	0,491
HS_3Reiche	3*P1; 15*P2; 15*P3	186	159	1,17	A	0,059	0,412
3Reiche	3*P1; 12*P2; 12*P3	153	135	1,33	A	0,061	0,415
3ReicheMinus1€	3*P1; 12*P2; 12*P3	153	111	1,38	A	0,061	0,625

Zur Operationalisierung der Hypothesen 1a, 1b und 1c wurden die Faktoren Fairness, Effizienz und individuelle Auszahlung systematisch variiert¹². Aus dem Verhältnis der eigenen Auszahlung zur Durchschnittsauszahlung lassen sich entsprechend dem ERC-Modell von Bolton und Ockenfels (2000) insgesamt 42 Fairness-Terme berechnen. Für 12 der 54 Fälle ist ein Fairness-Term aufgrund negativer Auszahlungen nicht definiert. Hierbei stellen das Labortreatment Gleich und das Hörsaal-treatment HS_Gleich Extremfälle dar, bei denen die Auszahlungsmatrix bewusst so gewählt wurde, dass für P3 die Optionen A und B identische Werte im Fairnessterm besitzen – P3 erhält in beiden Optionen jeweils exakt den Mittelwert der Gruppengesamtauszahlung, sodass ein optimaler Grad an Fairness besteht¹³. Somit ist Option B gleich fair (H1a)

12 Das Disparitätsmaß Gini-Koeffizient misst die relative Konzentration (Berechnung ohne Berücksichtigung eines Zeilenindex).

13 Für das Experiment von Bolton und Ockenfels (2003) ergeben sich als Vergleichsmaßstab folgende Parameter: Effizienzquotient 0,87, effiziente Option B, Gini-Koeffizient Option A = 0,0 und Gini-Koeffizient Option B = 0,333. In der Untersuchung von Bolton und Ockenfels liegt der Quotient aus eigener und durchschnittlicher Auszahlung bei Option A bei 1 und bei Option B bei 1,13, so dass hier anders als in unserem Design das ERC-Modell nicht

bei höherer Effizienz (H1b) und höherem Individualeinkommen (H1c). Hieraus lässt sich aus dem ERC-Modell prognostizieren, dass keine Person P3 die Option A wählt.

Da das ERC-Modell somit für P3 nur die Wahl von Option B zulässt, ist für eigen-nützige Individuen ausschließlich das Entscheidungsmuster BAB prognostizierbar. Nur im theoretischen Ausnahmefall, dass ein Individuum überhaupt nicht an Eigennutz interessiert ist, sind die Verhaltensmuster AAB (extreme Ungleichheitsaversion) und BBB (extreme Ausprägung von Effizienz) theoretisch denkbar.

In den drei Treatments HS_3Reiche, 3Reiche und 3ReicheMinus1€ ist Option A nicht nur fairer, sondern auch effizienter als Option B. Bei HS_3Reiche und bei 3Reiche nehmen die Fairnessterme für P1 sehr hohe und für P2 sehr niedrige Werte an, auch die Differenz der Fairnessterme zwischen den Optionen A und B ist in diesen Treatments am größten (beispielsweise erhält im Treatment 3Reiche eine Person P1 in Option A 7€ [Mittelwert 5,67€, Quotient 1,24] und in Option B 13€ [Mittelwert 5,00€, Quotient 2,6]). Für alle Personenrollen gilt, dass Option A sowohl viel fairer (H1a) als auch effizienter (H1b) ist und zudem für P2 eine höhere individuellen Auszahlungssumme gewährleistet (H1c), so dass im Vergleich zu anderen Treatments ein starker Effekt in Richtung einer Wahl von Alternative A für alle drei Personenrollen prognostiziert wird.

Die Operationalisierung von Hypothesen der distributiven Gerechtigkeitsforschung stellt eine größere Herausforderung dar, da das Gerechtigkeitsmotiv mathematisch (beispielsweise als arithmetischer Mittelwert oder Gini-Koeffizient) nicht definiert ist. Zudem bleiben einschränkende Randbedingungen ex ante unklar. Jedoch lässt sich die grundsätzliche Aussage ableiten, dass in der vorliegenden Spielsituation die Bedingung von Hypothese H2 („Ressource, die vom Himmel fällt“) erfüllt ist und somit in der Mehrheitsentscheidung distributive Gleichverteilung im Vordergrund steht, also vermehrt Alternative A, aber auch intrapersonell häufiger das Entscheidungsmuster AAA gewählt wird. Fairness sollte zudem als Entscheidungsgrund genannt werden.

Die Do-no-harm-Heuristik wirkt zwar ebenfalls in Richtung der gleichverteilten Option A, jedoch wird dieser Effekt nicht schwächer, wenn nur wenige Personen von Ungleichheit betroffen sind. Der Sonderfall des Identifiable-Victim-Effekt tritt sogar ausschließlich im Falle einer kleinen benachteiligten Minderheit auf. Diese wird operationalisiert durch die Treatments mit drei „Armen“. Eine negative Auszahlung, operationalisiert durch die Treatments mit minus ein Euro Auszahlung für P2, erhöht die Sichtbarkeit und das Leid der Opfer. Der Effekt wird durch das Treatment 3ArmeMinus1€ operationalisiert. Als Gesamteffekt wird prognostiziert, dass der Identifiable-Victim-Effekt die Faktoren Effizienz, Fairness (in der Definition ökonomischer Modelle als Abweichung vom Mittelwert) und Eigennutz kompensiert bzw. überkompensiert, die in entgegengesetzter Richtung wirken. Hieraus leiten sich die folgenden drei konkreten Operationalisierungen für Hypothese H3 ab: (1) Spieler im Treatment 3ArmeMinus1€ votieren in höherem Ausmaß für Option A im

falsifizierbar ist, da die Wahl von Option A aus Fairnessaspekten nicht ausgeschlossen werden kann.

Vergleich zum Treatment 3Arme. (2) Spieler im Treatment 3ArmeMinus1€ votieren in gleichem oder in höherem Ausmaß für Option A im Vergleich zu den Treatments GleichverteiltMinus1€ und 3ReicheMinus1€, obwohl in diesen beiden Treatments mehr P2 unter der Auszahlung Minus 1€ leiden würden. (3) Fairness für P2 wird im Treatment 3ArmeMinus1€ häufiger als Grund genannt als in allen anderen Treatments.

Hypothese H4 wird anhand des Vergleichs zwischen Hörsaal und Labor operationalisiert¹⁴. Hierbei sollte im Labor der Frame „Zweckrational Geld verdienen“ stärker sein, in geringerem Ausmaß Reframing auftreten und somit grundsätzlich der Anteil auszahlungsmaximierender Entscheidungen höher liegen als im Hörsaal (H4).

Die Low-Cost-Hypothese (formalisiert innerhalb des MFS) wird wie folgt operationalisiert: In der Befragung sollten unabhängig von der Personenrolle soziale Einstellungen geäußert werden. Bezüglich des Verhaltens sind jedoch nur für P3 die Kosten zur Realisierung relativ niedrig und die Tendenz, wertrational zu entscheiden (Match und Nutzen von F_{WR}), hoch, so dass es hier zu einem Reframing und einem hohen Anteil nichteigennützigen Entscheidungsverhaltens kommt (H5). Hierbei werden intraindividuell konsistente Entscheidungsmuster prognostiziert – Kombinationen wie AAB oder BBA, also Fairness nur da zu wählen, wo sie am meisten kostet, dürften laut LCH von keinem Individuum gewählt werden.

In Spiel 2 soll die zeitliche und inhaltliche Konsistenz bezüglich der in Spiel 1 gewählten Entscheidungen überprüft werden anhand der Gruppe der Spieler, die in Spiel 1 AAA (Gleichverteilung) wählen, in deren Gesamtgruppe per Mehrheitsvotum die Alternative B gewählt wird und die als P1 ausgelost wurden. Die Spielergruppe mit dem Verhaltensmuster AAA wird betrachtet, weil hier die Diskrepanz zwischen offenkundiger individueller Präferenz für Gleichverteilung und tatsächlich realisierter Gruppenentscheidung für die ungleiche Auszahlungsoption am größten ist. Unter der Annahme, dass Ungleichheitsaversion ein zeitlich konstanter Persönlichkeitszug ist, sollten in Spiel 2 gerade von P1 mit der höchsten Auszahlung und gleichzeitig größten Ungleichheitsaversion erhebliche Transfers in Richtung der benachteiligten P2 erfolgen. Für Spiel 2 lässt sich die aus der Konsistenzannahme abgeleitete Hypothese H6 wie folgt operationalisieren: Wenn per Mehrheitsentscheid Option B in Spiel 1 gewählt wird, dann verteilen P1 mit dem fairen Entscheidungsmuster AAA ihr Einkommen von 13€ zu einem signifikanten Anteil im anschließenden Spiel 2 um und dieser Anteil ist größer als bei P1 mit anderen Entscheidungsmustern.

14 Da die Gruppengröße zwischen Hörsaal (33 Vpn) und Labor (27 Vpn) differiert, kann das Setting nur annäherungsweise als isolierter Effekt angenommen werden.

Durchführung

Die Hörsaalexperimente wurden als Paper-Pencil-Experiment, d.h. als schriftliche Befragung mit Papierfragebogen zu Beginn von zwei Vorlesungen bzw. Seminaren der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln mit Studierenden aus den Fächern Ökonomie, Sozialwissenschaften und Psychologie durchgeführt. Insgesamt wurden im Hörsaal Daten von 108 Studierenden erhoben. Sieben Studierende wurden von der Auswertung ausgeschlossen, da sie Verständnisfragen nicht richtig beantwortet hatten. In allen drei Treatments befanden sich mindestens 33 Spieler. Aus den verbliebenen 101 Studierenden wurden daher zwei per Los ausgeschlossen, so dass jeweils exakt 33 Versuchspersonen pro Treatment verblieben. Diese 99 Spieler erhielten die der ihnen zugewiesenen individuellen Personenrolle und der Mehrheitsentscheidung in ihrer Gruppe entsprechenden Beträge ausgezahlt. Die insgesamt neun übrigen Versuchspersonen erhielten die ihrer Gruppe und zugewiesenen Personenrolle (ohne Teilnahme an der Entscheidung) entsprechenden Beträge ausgezahlt. Die Zeitdauer der Hörsaalexperimente betrug etwa 15 Minuten. Eine Show-Up-Gebühr wurde nicht entrichtet. Im Durchschnitt wurden an die Versuchspersonen 7,27 € ausgezahlt.

Die Laborexperimente wurden im Kölner Laboratorium für Wirtschaftsforschung (CLER) in sechs Sessions mit je 27 Spielern durchgeführt, die über die Rekrutierungssoftware ORSEE (Greiner 2015) aus dem CLER-Teilnehmerpool ausgewählt wurden. Sowohl den Versuchspersonen als auch den Ersatzspielern wurde wie in der experimentellen Wirtschaftsforschung üblich eine Show-Up-Gebühr von 2,50€ ausgezahlt. Das Laborexperiment wurde mit z-Tree (Fischbacher 2007) programmiert. An den Experimenten nahmen insgesamt 162 Versuchspersonen teil. Alle Teilnehmer bis auf eine Person (Hotelangestellte) waren Studierende zumeist wirtschafts- und sozialwissenschaftlicher Studiengänge. Durchschnittlich erhielten die Versuchsteilnehmer am Ende der Session nach der Spendenoption (Spiel 2), inklusive der Show-Up-Gebühr, 9,03€. Zusätzlich wurden von den Teilnehmern 78,65€ an die Hilfsorganisation Ärzte ohne Grenzen gespendet.

Von den insgesamt 270 Versuchsteilnehmern wurden wie oben beschrieben $N = 261$ Personen berücksichtigt mit 783 Entscheidungen im Strategiemodus in Spiel 1 und 648 Entscheidungen in Spiel 2.

Die manuell erfassten und bereinigten Daten des Hörsaalexperiments wurden mit den über zTree in EXCEL erfassten Labordaten in einen gemeinsamen Datensatz integriert und mit der Software SPSS sowie zur Auswertung der qualitativen Daten nach Erstellung eines Kategorienschemas (Mayring 2010) mit der Software MAXQDA analysiert.

Ergebnisse

Im Folgenden werden die Ergebnisse zunächst deskriptiv dargestellt, um anschließend die einzelnen Hypothesen zu überprüfen. In Tabelle 3 sind die absoluten und relativen Häufigkeiten der Entscheidungen (im Strategiemodus) für Alternative A (relative Gleichverteilung) über alle 9 Treatments, sowie die jeweils per Voting von der Gesamtgruppe gewählte Alternative dargestellt.

Tabelle 3 Häufigkeiten Wahl von Alternative A (gleichmäßigere Verteilung) und Voting über alle Treatments

	Wahl von Alternative A: Anzahl (Prozent)						gewählte Alternative
Treatment	P1		P2		P3		
HS_3Arme	4	(12,1%)	31	(93,9%)	8	(24,2%)	B
3Arme	0	(0,0%)	26	(96,3%)	6	(22,2%)	B
3ArmeMinus1€	5	(18,5%)	26	(96,3%)	7	(25,9%)	B
HS_Gleich	3	(9,1%)	30	(90,9%)	6	(18,2%)	B
Gleich	3	(11,1%)	27	(100%)	5	(18,5%)	B
GleichMinus1€	4	(14,8%)	27	(100%)	8	(29,6%)	A
HS_3Reiche	9	(27,3%)	32	(97,0%)	19	(57,6%)	A
3Reiche	4	(14,8%)	26	(96,3%)	9	(33,3%)	A
3ReicheMinus1€	6	(22,2%)	27	(100%)	15	(55,6%)	A
Gesamtsumme (N=261)	38	(14,6%)	252	(96,6%)	83	(31,8%)	

Die von der Gesamtgruppe gewählte Alternative ergibt sich zwar aus den per Los zugeordneten Personenrollen – eine Simulation von Zufallsziehungen kommt jedoch mehrheitlich zum gleichen Gruppenergebnis. Die Erhebung der Entscheidungen im Strategiemodus bedingt eine Analyse intraindividuelle Entscheidungsmuster über alle drei Personenrollen. Dies ist in Tabelle 4 dargestellt.

Tabelle 4 Häufigkeiten Entscheidungsmuster über alle Treatments

Treatment	Entscheidungsmuster					Gesamtsumme
	AAA	AAB	BAA	BAB	BBB	
HS_3Arme	3 (9,1%)	1 (3,0%)	5 (15,2%)	22 (66,7%)	2 (6,1%)	33 (100%)
3Arme			6 (22,2%)	20 (74,1%)	1 (3,7%)	27 (100%)
3ArmeMinus1€	5 (18,5%)		2 (7,4%)	19 (70,4%)	1 (3,7%)	27 (100%)
HS_Gleich	3 (9,1%)		3 (9,1%)	24 (72,7%)	3 (9,1%)	33 (100%)
Gleich	2 (7,4%)	1 (3,7%)	3 (11,1%)	21 (77,8%)		27 (100%)
GleichMinus1€	4 (14,8%)		4 (14,8%)	19 (70,4%)		27 (100%)
HS_3Reiche	8 (24,2%)	1 (3,0%)	11 (33,3%)	12 (36,4%)	1 (3,0%)	33 (100%)
3Reiche	4 (14,8%)		5 (18,5%)	17 (63,0%)	1 (3,7%)	27 (100%)
3ReicheMinus1€	6 (22,2%)		9 (33,3%)	12 (44,4%)		27 (100%)
Gesamtsumme	35 (13,4%)	3 (1,1%)	48 (18,4%)	166 (63,6%)	9 (3,4%)	261 (100%)

Über alle Treatments erweist sich das auszahlungsmaximierende Muster BAB als dominant, gefolgt von AAB und AAA. Die Entscheidungsmuster ABA, ABB und BBA wurden von keinem Spieler gewählt.

Tabelle 5 Arithmetische Mittelwerte quantitative Fragen 6-stufige Skala
(1= stimme überhaupt nicht zu; 6 = stimme voll und ganz zu)

Item	3Arme	3Arme Minus1€	Treatment			
			Gleich	Gleich Minus1€	3Reiche	3Reiche Minus1€
möglichst hohe Geldauszahlung	5,30	5,26	5,26	4,44	5,11	4,48
Versuch faires Verhalten	3,30	3,04	3,26	3,85	3,15	4,04
Wahl von Effizienz wichtig	4,33	4,04	3,96	4,22	3,89	3,81
Alternative A ist fair	4,37	4,89	4,56	4,33	4,63	4,74
Alternative B ist fair	2,63	1,85	1,93	2,52	2,04	1,59
Alternative A ist effizient	3,44	3,81	4,07	3,96	4,00	4,41
Alternative B ist effizient	3,70	3,52	3,30	3,26	3,22	2,48

Im Anschluss an Spiel 1 wurden in den Labortreatments die Gründe für das gewählte Entscheidungsmuster und die Klassifizierung der Optionen anhand schriftlicher Survey-Fragen mittels sechsstufigen Ratingskalen erhoben, die in Tabelle 5 dargestellt sind.

Die ersten drei Items weisen eine starke Verhaltensnähe bezogen auf die Gesamtscheidungssituation auf, während die letzten vier Items die Einstellung gegenüber den Entscheidungsoptionen widerspiegeln. Eine möglichst hohe Geldauszahlung wird von den Versuchspersonen in allen Treatments als wichtigstes Entscheidungsmotiv genannt, aber auch Fairness und Effizienz werden als wichtige Verhaltensgründe angeführt. Während das Eigennutzmotiv mit dem oben dargestellten tatsächlichen Entscheidungsverhalten korrespondiert, scheinen sich Fairnessmotive weniger und Effizienz motive fast überhaupt nicht im Verhalten niederzuschlagen. Objektive Fairness- und Effizienzmaße entfalten kaum Einfluss auf die Bewertung der unterschiedlichen Treatments. Mit Ausnahme von 3Arme wird Alternative A in allen Treatments sowohl als fairer als auch als effizienter als Alternative B bewertet.

Die getrennt für jede Personenrolle P1, P2 und P3 erhobenen qualitativen Entscheidungsgründe lassen sich bezüglich vieler gemessener Aussagen nicht den oben dargestellten Theorien zuordnen, so dass eine rein deduktive Kategorienbildung nach Mayring (2010) ausscheidet. Die Aussagen werden daher in einem kombiniert induktiv-deduktiv gebildeten Kategoriensystem codiert. Die große Theorieferne vieler Aussagen wirft grundsätzliche Fragen auf, die in späteren Untersuchungen zu klären sind, so dass hier auf Subkategorien verzichtet wird und die qualitativen Aussagen lediglich zur Plausibilisierung der quantitativen Ergebnisse dienen. Die folgenden, als Dummy-Variablen codierten Kategorien werden häufig genannt: „Eigennutz/Geldmaximierung“, „Fairness für Person 2“, „Effizienz“, „Verlustvermeidung“, „Orientierung an den anderen Mitspielern“, „gleicher Auszahlungsbetrag bedeutet Fairness“, „weniger Geld ist genug“, „große Gruppe wurde wahrgenommen“, „große Differenz zwischen Auszahlungsbeträgen“, „Wahrscheinlichkeiten berechnet“. Über alle Treatments und Personenrollen wird Eigennutz/Geldmaximierung als wichtigster Entscheidungsgrund genannt. Im Folgenden werden die Ergebnisse zu den einzelnen Hypothesen dargestellt.

Hypothese H1a, H1b und H1c (Ungleichheitsaversion gemäß ökonomischer Entscheidungsmodelle)

Die in Tabelle 3 deskriptiv dargestellten Unterschiede zwischen den einzelnen Treatments erweisen sich statistisch lediglich für Personenrolle P3 als signifikant (Kruskal-Wallis- H -Test über die Labortreatments für P3 $H_{(5)} = 10,925$; $p = 0,053$ und für die Hörsaaltreatments $H_{(2)} = 13,229$; $p \leq 0,001$). Der Chi-Quadrat-Test zwischen Entscheidungsmustern und den in Skalen abgefragten Entscheidungsgründen ergibt über die sechs Labor-Treatments für Eigennutz („möglichst hohe Geldauszahlung“) ein $\chi^2_{20} = 73,257$; $p \leq 0,001$ mit einer mittleren Effektstärke von $V = 0,336$ und für Fairness

(„Versuch faires Verhalten“) 98,419; $p \leq 0,001$ mit einer mittleren Effektstärke von $V = 0,390$. Für den Entscheidungsgrund Effizienz („Wahl von Effizienz wichtig“) ergibt sich kein signifikanter Zusammenhang.

Für die qualitativ erhobenen Entscheidungsgründe zu den einzelnen Entscheidungen in der jeweiligen Personenrolle wird die Prädiktorkraft mittels logistischer Regression getrennt für Labor- und Hörsaalbedingung getestet. Für die Hypothesen H1a, H1b und H1c wurden „Eigennutz/Gewinnmaximierung“, „Effizienz“ und „gleicher Auszahlungsbetrag bedeutet Fairness“ ausgewählt. In den *Labortreatments* ergibt sich: Für die Entscheidung als P1 ist für Eigennutz der Wald-Test = 14,808; $df = 1$; $p \leq 0,001$ höchstsignifikant. Die Vorhersagekraft des Modells verbessert sich von 86,4% auf 98,1% bei einem Nagelkerke- R^2 von 0,859. Für P2 ergibt sich kein signifikanter Zusammenhang. Für die Entscheidung als P3 liegen sowohl für die Variable Treatment (Wald-Test = 6,5; $df = 1$; $p = 0,011$), als auch für Eigennutz (Wald-Test = 35,126; $df = 1$; $p \leq 0,001$ höchstsignifikant) und für Fairness (Wald-Test = 7,819; $df = 1$; $p = 0,005$) hochsignifikante Ergebnisse vor bei einem Nagelkerke- R^2 von 0,784 und einer Verbesserung der Prognosekraft von 69,1% auf 90,1%. In den *Hörsaal-treatments* ergibt sich: Für die Entscheidung als P1 ist für Eigennutz der Wald-Test = 17,226; $df = 1$; $p \leq 0,001$ höchstsignifikant. Die Vorhersagekraft des Modells verbessert sich von 83,8% auf 97,0% bei einem Nagelkerke- R^2 von 0,826. Für P2 ergibt sich kein signifikanter Zusammenhang. Für P3 ist in der Kategorie Eigennutz der Wald-Test 19,038; $df = 1$; $p \leq 0,001$ höchstsignifikant bei einem Nagelkerke- R^2 von 0,812 und einer Verbesserung der Prognosekraft von 66,7% auf 91,9%. Im Folgenden werden die konkreten operational formulierten Hypothesen getestet.

Die Zahl der P3, die in den Treatments Gleich und HS_Gleich die Option A wählt, liegt niedriger als in den anderen Treatments. In HS_Gleich wählen sechs (18,2%) P3 und im Labortreatment Gleich wählen fünf (18,5%) P3 die Alternative A. Die Prognose, dass P3 in diesen Treatments nicht Alternative A wählt, wird nicht bestätigt. Auch die Prognose bezüglich des Entscheidungsmusters (ausschließlich BAB, im Extremfall AAB oder BBB) kann nicht bestätigt werden: In diesen beiden Treatments zeigt sich mehr als in den anderen Treatments das Entscheidungsmuster BAB (72,7% in HS_Gleich und 77,8% in Gleich). Aber selbst unter Hinzunahme der Extrem-Entscheidungsmuster BBB (drei Personen bzw. 9,1% im Treatment HS_Gleich) und AAB (eine Person [3,7%] im Treatment Gleich) kann die Prognose bezüglich der Entscheidungsmuster nicht gestützt werden.

In den drei Treatments HS_3Reiche, 3Reiche und 3ReicheMinus1€ mit wenigen „Reichen“ wird als Gruppenentscheid die gleicher verteilte Alternative A gewählt. Deskriptiv votieren in den beiden Treatments HS_3Reiche und 3ReicheMinus1€ mehr Personen P1 und P3 für die gleicher verteilte Option A als in allen anderen Treatments. In allen drei Treatments unterscheidet sich das Entscheidungsverhalten in der Subgruppe P2 kaum vom Entscheidungsverhalten in den übrigen sechs Treatments. Die punktbiseriale Korrelation zwischen dem Entscheidungsverhalten von P1 und P3 in den einzelnen Treatments und dem in den qualitativen Fragen geäußerten Entscheidungsgrund Fairness ist für alle neun Treatments (einschließlich aller drei Treatments mit wenigen „Reichen“) signifikant, mit Ausnahme 3Arme

(P1) und GleichMinus (P3), für die jeweils keine Signifikanz besteht. Der Faktor Effizienz ist bezogen auf die drei Treatments HS_3Reiche, 3Reiche und 3ReicheMinus1€ nur signifikant für das Treatment 3Reiche (für P1 ist die Korrelation mit einem Pearson-Korrelationskoeffizient $r = -0,470$ signifikant mit $p = 0,013$ [2-seitig]; für P3 ist die Korrelation signifikant mit $p = 0,039$ [2-seitig] und $r = -0,400$). Von diesen drei Treatments wird lediglich für das Treatment 3Reiche die Alternative A signifikant von Person 3 als effizient wahrgenommen ($r = -0,503$ mit $p = 0,008$ [2-seitig]). Jedoch wird im gleichen Treatment auch von Personen 3 die Alternative B als effizient wahrgenommen ($r = 0,488$; $p = 0,010$ [2-seitig]). Der Faktor Eigennutz (Hypothese H1c) dominiert Entscheidungen und Entscheidungsgründe. Fairness (Hypothese H1a) und vor allem Effizienz (Hypothese H1b) spielen im Verhalten, in den Verhaltensmustern und in den geäußerten Gründen nur eine untergeordnete Rolle, jedoch ist in den drei Treatments mit wenigen „Reichen“ die Veränderung von Effizienz- und Fairnessmaßen so dominant, dass es zu einem spürbaren Sinken des eigennützigen Verhaltensmusters BAB kommt. In den Verhaltensgründen finden sich nicht die in den Hypothesen H1a, H1b und H1c postulierten Zusammenhänge.

Hypothese H2 (Distributive Gerechtigkeitsnorm)

Als P2 wählen 96,7% der Versuchspersonen Alternative A. In der Subgruppe P1 wird von 14,4% Alternative A gewählt und in der Subgruppe P3 von 31,7%. Für das Entscheidungsmuster AAA votieren 13,4% mit den Extremwerten 24,2% im Treatment HS_3Reiche und 9,1% in HS_3Arme. Für die qualitativ erhobenen Entscheidungsgründe zu den einzelnen Entscheidungen in der jeweiligen Personenrolle wird für H2 die Prädiktorkraft des Faktors „gleicher Auszahlungsbetrag bedeutet Fairness“ mittels logistischer Regression getestet. Für die *Labortreatments* ergibt sich: Für P3 ist der Wald-Test = 14,214; $df = 1$; $p \leq 0,001$ höchstsignifikant bei einem Nagelkerke- $R^2 = 0,272$ und der Verbesserung der Prognosekraft von 69,1% auf 78,4%. Für die beiden anderen Personenrollen ergibt sich kein signifikanter Zusammenhang. Für die *Hörsaal treatments* ergibt sich: Für die Entscheidung als P1 ist der Wald-Test = 15,385; $df = 1$; $p \leq 0,001$ höchstsignifikant. Die Vorhersagekraft des Modells verbessert sich von 83,8% auf 90,9% bei einem Nagelkerke- R^2 von 0,411. Für die beiden anderen Personenrollen ergibt sich kein signifikanter Zusammenhang. Hypothese H2 kann nicht bestätigt werden.

Hypothese H3 (Identifiable-Victim-Effekt)

Der Mann-Whitney-*U*-Test durchgeführt für 3Arme versus 3ArmeMinus1€ ist lediglich für P1 in 3Arme versus 3ArmeMinus1€ hochsignifikant ($U = 297$; die Rangsumme von 3ArmeMinus1€ ($\Sigma_{3AM} = 675$) liegt unter der Rangsumme von 3Arme ($\Sigma_{3A} = 810$); $p = 0,01$). Die Zusammenhänge für 3ArmeMinus1€ versus 3ReicheMinus1€ bzw. GleichMinus1€ sind nicht signifikant. Dieser Befund stützt den Identifiable-Victim-Effekt.

Für die qualitativ erhobenen Entscheidungsgründe zu den einzelnen Entscheidungen in der jeweiligen Personenrolle wird die Prädiktorkraft mittels logistischer Regression getestet. Für die Hypothese H3 wurden die von den Versuchspersonen genannten Gründe bezüglich „Verlustvermeidung“ und bezüglich „Fairness für Person 2“ ausgewählt. Für die *Labortreatments* folgt: Für die Entscheidung als P3 ergibt der Wald-Test = 7,889; $df = 1$; $p = 0,005$ für Verlustvermeidung. Die andere Kategorie ist nicht signifikant. Die Vorhersagekraft des Modells verbessert sich von 69,1% auf 85,2% bei einem Nagelkerke- R^2 von 0,521. Für die beiden anderen Personenrollen P1 und P2 sind die Zusammenhänge nicht signifikant. Für die *Hörsaalbehandlungen* folgt: Für die Entscheidung als P1 ergibt der Wald-Test = 5,190; $df = 1$; $p = 0,023$ für Verlustvermeidung. Die andere Kategorie ist nicht signifikant. Die Vorhersagekraft des Modells verbessert sich von 83,8% auf 89,9% bei einem Nagelkerke- R^2 von 0,381. Für die beiden anderen Personenrollen P2 und P3 ergeben sich keine signifikanten Zusammenhänge.

In der punktbiserialen Korrelation besteht für 3ArmeMinus1€ für P1 und P3 ein signifikanter Zusammenhang zwischen Entscheidungsverhalten und dem Skalitem „Alternative A ist fair“ ($P1 \rightarrow r = -0,383$; $p = 0,048$ und $P3 \rightarrow r = -0,356$; $p = 0,068$ (2-seitig)). Die Befunde zu den Entscheidungsgründen stützen den Identifiable-Victim-Effekt für das Treatment 3ArmeMinus1€.

Hypothese 4 (Hörsaal versus Labor)

Zwischen den Treatments 3Arme versus HS_3Arme, Gleich versus HS_Gleich sowie 3Reiche versus HS_3Reiche kann jeweils kein signifikanter Unterschied des Entscheidungsverhaltens zwischen den Settings gefunden werden. Im Treatment 3Reiche ergibt sich bei einem $U = -1,857$ und $p < 0,05$ ein signifikanter Unterschied für P3 zwischen Labor und Hörsaal. Für die beiden anderen Personengruppen P1 und P2 ergeben sich keine signifikanten Ergebnisse. In der Bedingung 3Arme versus HS_3Arme kommt es lediglich für P1 zu einem signifikanten Effekt. P2 und P3 zeigen keine signifikanten Unterschiede. Die Hypothese H4, dass der Frame F_{ZG} in der Laborbedingung stärker ist als im Hörsaal, kann nicht bestätigt werden.

Hypothese H5 (Low-Cost)

Intrapersonell besteht zwischen den Entscheidungen von P1 mit relativ hohen Kosten und P3 mit relativ geringen Kosten für die Realisierung einer als gerecht empfundenen Alternative in allen Treatments deskriptiv ein Unterschied. Die größte Differenz zwischen P1 und P3 zeigt sich bei den Treatments 3ReicheMinus1€ (33,4% Spannbreite), HS_3Reiche (30,3%) und 3Arme (22,2%). Zudem wird die laut LCH unzulässige Kombinationen AAB jeweils von einer Vpn in den Treatments HS_3Arme, Gleich und HS_3Reiche gewählt. Im Einklang mit der LCH wird das Entscheidungsmuster BBA nicht gewählt. Der Wilcoxon-Test über Unterschiede im intrapersonel-

len Entscheidungsverhalten zwischen P1 und P3 ist für alle Treatments signifikant außer für die Treatments 3ArmeMinus1€ und GleichverteiltPlus.

Für die qualitativ erhobenen Entscheidungsgründe zu den einzelnen Entscheidungen in der jeweiligen Personenrolle wird die Prädiktormöglichkeit der Kategorie „weniger Geld ist genug“ mittels logistischer Regression getestet. Für die *Labortreatments* ergibt sich: Für P3 ist der Wald-Test = 16,701; $df = 1$; $p \leq 0,001$ höchstsignifikant bei einem Nagelkerke- $R^2 = 0,148$. Die Vorhersagekraft hat sich für P3 von 69,1% auf 74,7% verbessert. Für die *Hörsaal treatments* ergibt sich: Hier ist für P3 der Wald-Test = 16,685; $df = 1$; $p \leq 0,001$ höchstsignifikant mit einem Nagelkerke- $R^2 = 0,471$ und der Verbesserung der Vorhersagekraft von 66,7% auf 83,8%. Für die anderen beiden Personenrollen P1 und P2 bestehen keine signifikanten Zusammenhänge. Trotz eines geringen „Rauschens“ (unlogische Entscheidungsmuster) bewährt sich H5 insgesamt.

Spiel 2: Intersituative Konsistenz des Entscheidungsverhaltens (H6)

Die in Spiel 1 erhaltene Geldsumme wurde wie folgt (Tabelle 6) an die Subgruppen P1, P2 und P3 sowie an die Hilfsorganisation Ärzte ohne Grenzen realloziert.

Tabelle 6 Umverteilung in Spiel 2, Beträge jeweils in €

Treatment	Gewählte Alternative	Geldumverteilung an P1	Geldumverteilung an P2	Geldumverteilung an P3	Geldumverteilung an Hilfsorganisation Ärzte ohne Grenzen	Gesamt	Gesamt anteilig zum Gruppeneinkommen (in Prozent)
3Arme	B	17,90	7,60	3,75	30,20	59,45	24,47
3ArmeMinus1€	B	2,50	14,00	2,50	10,30	29,30	12,36
Gleich	B	7,60	8,50	1,10	6,70	23,90	12,65
GleichMinus1€	A	2,00	5,71	2,55	9,04	19,30	11,91
3Reiche	A		0,94		6,16	7,10	4,64
3ReicheMinus1€	A	0,37	1,25	0,48	16,25	18,35	11,99
Gesamt		30,37	38,00	10,38	78,65	157,40	

Sowohl die Transfers an P1 als auch die starken Unterschiede zwischen den einzelnen Treatments beruhen im Wesentlichen auf extrem hohen Geldtransfers einzelner Versuchspersonen. Nur die 14€ im Treatment 3ArmeMinus1€ setzen sich aus den Reallokationen vieler Spieler (12 der 27 Versuchspersonen) zusammen. Zufällig sind alle drei Versuchspersonen, die das Kriterium von Hypothese H6 (fares Entscheidungsmuster AAA, Mehrheitsvoting für Option B, gelost als P1 mit 13€ Auszahlung) erfüllten, Gruppenmitglieder des Treatments 3ArmeMinus1€. Eine dieser drei Versuchspersonen behält die gesamten 13€, die zweite Vpn transferiert 1€ an P2 und

behält 12€, die dritte Vpn transferiert 1€ an P2, 2€ an ÄoG und behält 10€, so dass H6 bezüglich Transfers der zuvor in Spiel 1 geäußerten Verteilungspräferenz nicht bestätigt werden kann. Diese an Fairness orientierten P1 transferieren durchschnittlich 0,67€ (5,1% ihrer 13€ Gesamteinkommen) an P2, während die neun übrigen P1 demgegenüber im gleichen Treatment 3ArmeMinus1€ einen höheren durchschnittlichen Geldbetrag von 0,91€ (9,1% von 13€) an P2 transferieren. Auch hier kann H6 nicht gestützt werden. Lediglich das Spendenvolumen der drei fairen P1 an ÄoG ist mit durchschnittlich 0,67€ höher als die durchschnittlich 0,42€ Spenden der übrigen neun Versuchspersonen. Jedoch nennt von den drei P1 mit dem Entscheidungsmuster AAA nur eine Vpn Gleichverteilung des Geldes in Spiel 1 als Verhaltensgrund, während die beiden anderen Versuchspersonen Eigennutz als Grund für die Wahl von AAA angeben, um nicht bei der Mehrheitsentscheidung und bei einer Zulassung in die Subgruppe P2 nur Minus 1€ zu erhalten. Ob diesen beiden Versuchspersonen die Untauglichkeit ihres Verhaltens für ihre Zielerreichung bewusst war, geht aus den gemessenen Daten nicht hervor.

Diskussion, Fazit und Ausblick

Ein zentraler Befund der Untersuchung ist die überragende Bedeutung von Eigennutz für das Entscheidungsverhalten in Gruppen, wie dies von Olsons Theorie kollektiven Handelns prognostiziert wird. Jedoch werden auch entgegengesetzte Konzepte gestützt, beispielsweise die Low-Cost-Hypothese bezüglich fairen Verhaltens. Viele Akteure wählen nur dann eine gleicher verteilte Option, wenn diese wenig kostet. Hier bleibt vor dem Hintergrund der empirischen Evidenz zu High Stakes eine offene Forschungsfrage, ob dies aus relativ und/oder absolut niedrigen Kosten resultiert). Wird jedoch nur eine sichtbare kleine Minderheit stark benachteiligt (Versuchsbedingung 3ArmeMinus1€), dann tritt der Identifiable-Victim-Effekt als Spezialfall der Do-no-harm-Heuristik in den Vordergrund, verdrängt das Equality-Gerechtigkeitsprinzip und wirkt dem Eigennutzprinzip entgegen. Der Identifiable-Victim-Effekt erweist sich zudem auch in Spiel 2 als robust.

In allen neun Treatments erweist sich in den intraindividuellen Mustern des Entscheidungsverhaltens und in den quantitativ und qualitativ geäußerten Verhaltensgründen als robust, dass mentale Prozesse und Entscheidungsverhalten von benachteiligten P2 nicht von Fairness (wie vom ERC-Modell und sozialpsychologischen Equality-Prinzip postuliert), sondern von purer Eigennutzmaximierung im Sinne des Konzepts der Ungerechtigkeitssensibilität dominiert werden.

Ökonomische Konzepte der Ungleichheitsaversion wie das ERC-Modell im Speziellen und passive Wahrnehmungstheorien im Allgemeinen können nicht bestätigt werden. Die aus dem Güth-van Damme-Spiel abgeleitete Modellierung, dass die Struktur der Einkommensverteilung zwischen den übrigen Gruppenmitgliedern nicht relevant ist, muss zurückgewiesen werden. Insbesondere auch der von Bolton und Ockenfels (2003) postulierte Faktor Effizienz lässt sich weder in den

Verhaltensmustern replizieren, noch in den geäußerten Verhaltensgründen nachweisen.

Auch Konzepte der sozialpsychologischen Gerechtigkeitsforschung liefern keine validen Prognosen, solange ihre Einbettung in Entscheidungsprozesse nicht expliziert wird. Ohne Modellierung der Randbedingungen in Bezug auf die Frage, welches Gerechtigkeitsprinzip bei Kollision mit alternativen Prinzipien ausgewählt wird (z.B. Ungleichheitsaversion versus Verfahrensgerechtigkeit, Ungerechtigkeits-sensibilität, Effizienz oder Do-no-harm-Heuristik), sind ex ante keine konkreten Hypothesen ableitbar.

Zudem finden sich explorativ weder Hinweise auf einen isolierten Gruppengrößeneffekt, der in einer zukünftigen Untersuchung anhand einer systematischen Variation der Gruppengröße gezielt untersucht werden sollte, noch auf einen isolierten Ankereffekt negativer Auszahlungen, noch auf einen Einfluss des Settings (Hörsaal versus Labor), das ebenfalls anhand einer isolierten Variation von Ort und Gruppenzusammensetzung zu überprüfen ist. Das individuelle Spendenverhalten zugunsten der Mitspieler und der Hilfsorganisation Ärzte ohne Grenzen (Spiel 2) bestätigt die These des MFS von der Abgeschlossenheit der Spielsituationen – die Annahme situationsüberspannender sozialer Präferenzen kann hier nicht bestätigt werden.

Die Kombination und Integration von Methoden und Konzepten aus unterschiedlichen Fachdisziplinen erweist sich als tragfähig. Hierbei könnte sich das MFS als leistungsfähiges Instrument erweisen, inhaltliche Konzepte beispielsweise der sozialpsychologischen Gerechtigkeitsforschung in die Modellierung von Entscheidungsprozessen zu integrieren und Untersuchungsergebnisse in einem einheitlichen theoretischen Rahmen zu erklären. Aus Perspektive des MFS lassen sich die Untersuchungsergebnisse so interpretieren, dass neben rational kalkulierenden Eigennutzmaximierern und wertrationalen Akteuren, die immun gegen Störreize sind und unzerbrüchlich die Gleichverteilung wählen, einige Individuen auch nach dem Prinzip der „Angemessenheit“ heuristisch oder rational das passende Gerechtigkeitsprinzip auswählen. Das kollektive Verhalten konvergiert nicht in Richtung Eigennutz, sondern resultiert aus der Aggregation dieser drei Typen.

Aufgrund der geringen Zahl unabhängiger Beobachtungen und mangelnder Empirie in Bezug auf die Frage, ob sich die Erkenntnisse auch in weiteren Designvarianten als robust erweisen, sind diese Ergebnisse und Schlussfolgerungen möglicherweise nur auf einen sehr eingeschränkten Geltungsbereich begrenzt, so dass eine Verallgemeinerung der Schlussfolgerungen nicht möglich ist. Zudem bleibt die externe Validität und die Übertragbarkeit auf Phänomene in der Realität unklar, da aus Sicht der Autoren übliche Plausibilitätsüberlegungen zunächst durch die empirische Erhebung der Wahrnehmungs- und mentalen Entscheidungsprozesse in zukünftigen Untersuchungen abgesichert werden sollten. Weitere Grenzen der vorliegenden Arbeit ergeben sich aus dem Untersuchungszweck, durch eine Kombination von Methoden und Konzepten unterschiedlicher Disziplinen die Ergebnisse in eine multidisziplinäre Sicht zu integrieren, sodass die methodische und inhaltliche Breite zulasten einer geringeren Tiefe geht. Für eine Anwendung des MFS im öko-

nomischen Kontext könnte es sich möglicherweise als fruchtbar erweisen, geeignete normierte Instrumente zur Erhebung gedanklicher Modelle zu entwickeln.

Literatur

- Adams, J.Stacy. 1965. Inequity in Social Exchange. *Advances in Experimental Social Psychology* 2: 267-299.
- Ajzen, Icek, und James Sexton. 1999. Depth of processing, belief congruence, and attitude-behavior correspondence. In *Dual-process theories in social psychology*, Hrsg. Shelly Chaiken und Yaacov Trope, 117-138. New York, NY: Guilford Press.
- Andreoni, James. 1989. Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy* 97(6): 1447-1458. doi:10.1086/261662.
- Andreoni, James, und John Miller. 2002. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* 70(2): 737-753.
- Bar-Hillel, Maya, und Menahem Yaari. 1993. Judgments of distributive justice. Psychological perspectives on justice: Theory and applications. In *Psychological perspectives on justice: Theory and applications*, Hrsg. Barbara A. Mellers und Jonathan Baron, 55-84. Cambridge, MA: Cambridge University Press.
- Baron, Jonathan. 1995. Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making* 8(2): 71-83.
- Beckert, Jens, und Wolfgang Streeck. 2008. *Economic sociology and political economy: A programmatic perspective* (MPIfG Working Paper 08/4). Köln: Max-Planck-Institut für Gesellschaftsforschung.
- Best, Henning. 2009. Kommt erst das Fressen und dann die Moral? Eine feldexperimentelle Überprüfung der Low-Cost-Hypothese und des Modells der Frame-Selektion. *Zeitschrift für Soziologie* 38: 131-151.
- Best, Henning, und Clemens Kroneberg. 2012. Die Low-Cost-Hypothese. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 64(3): 535-561.
- Biniossek, Claudia, Dirk Betz, Detlef Fetchenhauer, und Markus Lünen. 2007. Fairness ohne Eigennutz. Handlungsannahmen in der ökonomischen Theorie und Auswirkungen auf Reformoptionen im Gesundheitswesen. *Sozialer Fortschritt* 9-10: 252-258.
- Biniossek, Claudia, und Detlef Fetchenhauer. 2007. Fairness als Kollektivgut. *Zur Psychologie des Wohlfahrtsstaats: Sonderheft der Zeitschrift für Wirtschaftspsychologie* 9(4): 68-82.
- Blanken, Irene, Niels van de Ven und Marcel Zeelenberg. 2015. A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, 41(4): 540-558.
- Blanken, Irene, Niels van de Ven, Marcel Zeelenberg und Marijn Meijers. 2014. Three attempts to replicate the moral licensing effect. *Social Psychology* 45(3): 232-238. doi: 10.1027/1864-9335/a000189
- Bolton, Gary E., und Axel Ockenfels. 2000. ERC: A Theory of equity, reciprocity, and competition. *The American Economic Review* 90(1): 166-193.

- Bolton, Gary E., und Axel Ockenfels. 2003. *The behavioral tradeoff between efficiency and equity when a majority rules* (#12-2003). Jena: Max Planck Institut.
- Brandts, Jordi, und Gary Charness. 2011. The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics* 14(3): 375-398.
- Braun, Norman, und Axel Franzen. 1995. Umweltverhalten und Rationalität. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47(2): 231-248.
- Brosig, Jeannette, Thomas Riechmann, und Joachim Weimann. 2007. *Selfish in the End? An Investigation of Consistency and Stability of Individual Behavior*. FEMM Working Paper No. 05. Magdeburg: Otto-von-Guericke-Universität Magdeburg.
- Camerer, Colin, und Richard H. Thaler. 1995. Anomalies: Ultimatums, dictators and managers. *The Journal of Economic Perspectives* 9(2): 209-219.
- Charness, Gary, und Matthew Rabin. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117(3): 817-869.
- Deutsch, Morton. 1975. Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues* 31(3): 137-149.
- Diekmann, Andreas. 1996. Homo ÖKOnomicus. Anwendungen und Probleme der Theorie rationalen Handelns im Umweltbereich. *Umweltsoziologie. Sonderheft* 36: 89-118.
- Diekmann, Andreas, und Peter Preisendörfer. 1992. Persönliches Umweltverhalten: Diskrepanzen zwischen Anspruch und Wirklichkeit. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 44(2): 226-251.
- Diekmann, Andreas, und Peter Preisendörfer. 1998. Umweltbewußtsein und Umweltverhalten in Low-und High-Cost-Situationen: Eine empirische Überprüfung der Low-Cost-Hypothese. *Zeitschrift für Soziologie* 27(6): 438-453.
- Diekmann, Andreas, und Thomas Voss. 2004. Die Theorie rationalen Handelns. Stand und Perspektiven. In *Rational-Choice Theorie in den Sozialwissenschaften. Anwendungen und Probleme*, Hrsg. Andreas Diekmann und Thomas Voss, 13-29. München: Oldenbourg.
- Engelmann, Dirk, und Martin Strobel. 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review* 94(4): 857-869.
- Esser, Hartmut. 1996. Die Definition der Situation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48(1): 1-34.
- Esser, Hartmut. 2001. Soziologie. Spezielle Grundlagen. Band 6: Sinn und Kultur. Frankfurt am Main: Campus.
- Esser, Hartmut. 2002a. In guten wie in schlechten Tagen? Das Framing der Ehe und das Risiko zur Scheidung. Eine Anwendung und ein Test des Modells der Frame-Selektion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54: 27-63.
- Esser, Hartmut. 2002b. Ehekrisen: Das (Re-)Framing der Ehe und der Anstieg der Scheidungsraten. *Zeitschrift für Soziologie* 31: 472-496.
- Esser, Hartmut. 2003. Der Sinn der Modelle. Antwort auf Götz Rohwer. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 55: 359-368.
- Esser, Hartmut. 2010. Das Modell der Frame-Selektion. Eine allgemeine Handlungstheorie für die Sozialwissenschaften. Soziologische Theorie kontrovers. *Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie* 50: 45-61.

- Esser, Hartmut, und Clemens Kroneberg. 2015. An integrative theory of action. The model of frame selection. In *Order on the Edge of Chaos*, Hrsg. Edward J. Lawler, Shane R. Thye und Jeongkoo Yoon, 63–85. New York, NY: Cambridge University Press.
- Erev, Ido, Eyal Ert, Alvin E. Roth, Ernan Haruvy, Stefan Herzog, Robin Hau, Ralph Hertwig, Terrence Stewart, Robert West und Christian Lebiere. 2009. A choice prediction competition: choices from experience and from description. *Journal of Behavioral Decision Making* 23(1): 15–47.
- Falk, Armin, und Urs Fischbacher. 2006. A theory of reciprocity. *Games and Economic Behavior* 54(2): 293–315.
- Fehr, Ernst, und Klaus M. Schmidt. 1999. A Theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114(3): 817–868.
- Fehr, Ernst, und Klaus M. Schmidt. 2010. On inequity aversion: A reply to Binmore and Shaked. *Journal of Economic Behavior & Organization* 73(1): 101–108.
- Festinger, Leon, Henry W. Riecken und Stanley Schachter. 1956. *When prophecy fails: A social and psychological study of a modern group that predicted the end of the world*. Minneapolis: University of Minnesota Press.
- Fetchenhauer, Detlef, Dominik H. Enste und Vanessa Köneke. 2010. *Fairness oder Effizienz*. München: Roman-Herzog-Institut.
- Fetchenhauer, Detlef, und Xu Huang. 2004. Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences* 36(5): 1015–1029.
- Fischbacher, Urs. 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10(2): 171–178.
- Forsythe, Robert, Joel L. Horowitz, Nathan E. Savin, und Martin Sefton. 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior* 6(3): 347–369.
- Granovetter, Mark. 1978. Threshold models of collective behavior. *American Journal of Sociology* 83(6): 1420–1443.
- Greiner, B. 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association* 1(1): 114–125. doi: 10.1007/s40881-015-0004-4.
- Güth, Werner, und Eric van Damme. 1998. Information, strategic behavior, and fairness in ultimatum bargaining: An experimental study. *Journal of mathematical Psychology* 42(2): 227–247.
- Isaac, R. Mark, James M. Walker, und Arlington W. Williams. 1994. Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups. *Journal of Public Economics* 54(1): 1–36.
- Jenni, Karen, und George Loewenstein. 1997. Explaining the identifiable victim effect. *Journal of Risk and Uncertainty* 14(3): 235–257.
- Kandel, Eric R., James Schwartz, und Thomas Jessell (Hrsg.). (2012). *Neurowissenschaften: Eine Einführung*. Heidelberg: Springer Spektrum.
- Karakayali, Serhat, und J. Olaf Kleist. 2016. EFA-Studie. Strukturen und Motive der ehrenamtlichen Flüchtlingshilfe (EFA) in Deutschland. Berlin: Humboldt-Universität zu Berlin.

- Kittel, Bernhard. 2015. Experimente in der Wirtschaftssoziologie: Ein Widerspruch? In *Experimente in den Sozialwissenschaften, Soziale Welt, Sonderband 22*, Hrsg. Marc Keuschnigg, und Tobias Wolbring, 79-104. Baden-Baden: Nomos.
- Klößner, Jennifer. 2016. Forschungsstand empirischer Studien zu freiwilliger Arbeit. In *Freiwillige Arbeit in gemeinnützigen Vereinen*, 105-177. Wiesbaden: Springer Fachmedien Wiesbaden.
- Kravitz, David. A., und Barbara Martin. 1986. Ringelmann rediscovered: The original article. *Journal of Personality and Social Psychology* 50(5): 936-931.
- Kroneberg, Clemens. 2005. Die Definition der Situation und die variable Rationalität der Akteure: Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie* 34(5): 344-363.
- Kroneberg, Clemens. 2011. *Die Erklärung sozialen Handelns: Grundlagen und Anwendung einer integrativen Theorie*. Wiesbaden: Springer-Verlag.
- Kroneberg, Clemens. 2014. Frames, scripts, and variable rationality: An integrative theory of action. In *Analytical Sociology. Actions and Networks*, Hrsg. Gianluca Manzo, 97-123. Chichester, GB: Wiley.
- Kühnel, Steffen, und Sebastian Bamberg. 1998a. Überzeugungssysteme in einem zweistufigen Modell rationaler Handlungen: Das Beispiel umweltgerechteren Verkehrsverhaltens. *Zeitschrift für Soziologie* 27(4): 256-270.
- Kühnel, Steffen, und Sebastian Bamberg. 1998b. Ist die Low-Cost Hypothese noch zu retten? Erwiderung auf Diekmann und Preisendörfer. *Zeitschrift für Soziologie* 27(4): 273-275.
- Liebe, Ulf, Elias Naumann und Andreas Tutić. 2017. Sozialer Status und prosoziales Handeln: Ein Quasi-Experiment im Krankenhaus. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 69(1): 109-129.
- Lindenberg, Siegwart, und Linda Steg. 2013. Goal-framing theory and norm-guided environmental behavior. In *Encouraging sustainable behavior*, Hrsg. Hans C.M. van Trijp, 37-54. New York, NY: Psychology Press.
- Lotz, Sebastian, und Detlef Fetchenhauer. 2012. Lifting the veil of ignorance in distributive justice – Evidence from a welfare state game. *Wirtschaftspsychologie* 14: 74-80.
- Lotz, Sebastian, Thomas Schlösser, Daylian M. Cain und Detlef Fetchenhauer. 2013. The (in)stability of social preferences: Using justice sensitivity to predict when altruism collapses. *Journal of Economic Behavior & Organization* 93: 141-148. doi: 10.1016/j.jebo.2013.07.012
- Mayring, Philipp. 2010. Qualitative Inhaltsanalyse. In *Handbuch Qualitative Forschung in der Psychologie*, Hrsg. Günter Mey und Katja Mruck, 601-613. Wiesbaden: VS Verlag für Sozialwissenschaften, doi: 10.1007/978-3-531-92052-8_42
- Mikula, Gerold. 1993. On the experience of injustice. *European Review of Social Psychology* 4(1): 223-244.
- Mitzkewitz, Michael, und Rosemarie Nagel. 1993. Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory* 22(2): 171-198.
- Musgrave, Richard Abel, und Peggy B. Musgrave. 5. Sub Aufl. 1989. *Public finance in theory and practice*. New York, NY: McGraw-Hill College.
- Olson, Mancur. 1965, 20. Aufl. 2002. *The logic of collective action. Public Goods and the theory of groups*. Cambridge, MA: Harvard University Press.

- Quandt, Markus, und Dieter Ohr. 2004. Worum geht es, wenn es um nichts geht?. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 56(4): 683-707.
- Rapoport, Amnon. 1988. Provision of step-level public goods: Effects of inequality in resources. *Journal of Personality and Social Psychology* 54(3): 432-440. doi: 10.1037/0022-3514.54.3.432
- Rapoport, Amnon, und Dalit Eshed-Levy. 1989. Provision of step-level public goods: Effects of greed and fear of being gyped. *Organizational Behavior and Human Decision Processes* 44(3): 325-344.
- Sachdeva, Sonya, Rumen Iliev und Douglas L. Medin. 2009. Sinning saints and saintly sinners the paradox of moral self-regulation. *Psychological Science* 20(4): 523-528.
- Sauermann, Jan. 2017. Do Individuals Value Distributional Fairness? How Inequality Affects Majority Decisions. *Political Behavior*. <https://doi.org/10.1007/s11109-017-9424-6>
- Schelling, Thomas C. 1968. The Life You Save May Be Your Own. In *Problems in Public Expenditure Analysis. Studies of Government Finance*, Hrsg. S. B. Chase Jr., 127-176. Washington, DC: The Brookings Institution.
- Schmitt, Manfred. 1990. *Konsistenz als Persönlichkeitseigenschaft? Moderatorvariablen in der Persönlichkeits- und Einstellungsforschung*. Berlin: Springer.
- Schmitt, Manfred. 1993a. *Abriß der Gerechtigkeitspsychologie*. Trier: Universität, Fachbereich I - Psychologie. <http://www.gerechtigkeitsforschung.de/berichte/beri070.pdf> (Zugegriffen: 01. Mai 2017)
- Schmitt, Manfred. 1993b. Handlung als Synthese von Person und Situation: Lehren aus der Konsistenzkontroverse. *Zeitschrift für Sozialpsychologie* 71-74.
- Schmitt, Manfred, Anna Baumert, Detlef Fetchenhauer, Mario Gollwitzer, Tobias Rothmund und Thomas Schlösser. 2009. Sensibilität für Ungerechtigkeit. *Psychologische Rundschau* 60(1): 8-22.
- Schmitt, Manfred, und Leo Montada. 1982. Determinanten erlebter Gerechtigkeit. *Zeitschrift für Sozialpsychologie* 13: 32-44.
- Schwarz, Norbert. 1985. Theorien konzeptgesteuerter Informationsverarbeitung in der Sozialpsychologie. *Theorien der Sozialpsychologie* 3: 269-291.
- Selten, Reinhard. 1967. Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopol-experiments. In *Beiträge zur experimentellen Wirtschaftsforschung*, Hrsg. Heinz Sauermann, 136-168. Tübingen: J.C.B. Mohr (Paul Siebeck).
- Selten, Reinhard. 1978. The chain store paradox. *Theory and Decision* 9(2): 127-159.
- Six, Bernd, und Thomas Eckes. 1996. Metaanalysen in der Einstellungs-Verhaltens-Forschung. *Zeitschrift für Sozialpsychologie* 27(1): 7-17.
- Smith, Vernon L. 1976. Experimental economics: Induced value theory. *The American Economic Review* 66(2): 274-279.
- Smith, Vernon L. 2010. Theory and experiment: What are the questions? *Journal of Economic Behavior* 73: 3-15. doi: 10.1016/j.jebo.2009.02.008
- Steiniger, Tim, Thomas Schlösser, Daniel Ehlebracht und Detlef Fetchenhauer. 2015. How justice sensitivity predicts equality preferences in simulated democratic systems. In *The different worlds of inequality: Psychological determinants and implications of economic*

- inequality*, Inauguraldissertation Tim Steiniger, 37-56. Köln: Universität zu Köln. <http://kups.ub.uni-koeln.de/6525/> (Zugegriffen: 01. Mai 2017)
- Stocké, Volker. 2002. *Framing und Rationalität: die Bedeutung der Informationsdarstellung für das Entscheidungsverhalten*. München: Oldenbourg.
- Streeck, Wolfgang. 2010. Does "Behavioural Economics" offer an alternative to neoclassical paradigm? *Socio-Economic Review* (Discussion Forum II: Behavioural Economics) 8(2): 387-397. doi: 10.1093/ser/mwq002
- Thaler, Richard H. 1999. Mental accounting matters. *Journal of Behavioral decision making* 12(3): 183-206.
- Tutic, Andreas. 2015. Warum denn eigentlich nicht? Zur Axiomatisierung soziologischer Handlungstheorie. *Zeitschrift für Soziologie* 44(2): 83-98.
- Tversky, Amos, und Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5(4): 297-323.
- Van Damme, Eric, Kenneth G. Binmore, Alvin E. Roth, Larry Samuelson, Eyal Winter, Gary E. Bolton, Axel Ockenfels, ... und Ofer H. Azar. 2014. How Werner Güth's ultimatum game shaped our understanding of social behavior. *Journal of Economic Behavior & Organization* 108: 292-318.
- Von Neumann, John, und Oskar Morgenstern. 1944. *Theory of games and economic behavior*. Princeton: Princeton University Press.
- Walster, Elaine, Ellen Berscheid und G. William Walster. 1973. New directions in equity research. *Journal of Personality and Social Psychology* 25(2): 151-176. doi: 10.1037/h0033967
- Walster, Elaine, Ellen Berscheid und William G. Walster. 1976. New directions in equity research. *Advances in Experimental Social Psychology* 9: 1-42.
- Weimann, Joachim, Jeannette Brosig-Koch, Timo Heinrich, Heike Hennig-Schmidt, Claudia Keser und Christian Stahr. 2014. *An Explanation of (First Round) Contributions in Public-Good Experiments*. CESifo Working Paper Series No. 5039.
- Wiswede, Günter 2012. *Einführung in die Wirtschaftspsychologie*. München: Reinhardt.

6 Fairness ohne Eigennutz

Handlungsannahmen in der ökonomischen Theorie und Auswirkungen auf Reformoptionen im Gesundheitswesen¹

Claudia Biniossek, Dirk Betz, Detlef Fetchenhauer & Markus Lünen

Zusammenfassung

Fairness aus Eigennutz (standardökonomische Verhaltenstheorie) und die Orientierung an Fairness der empfangenen Auszahlungen (Experimentelle Wirtschaftsforschung) werden in der ökonomischen Literatur als Ursachen für kooperatives Verhalten angesehen. In Kollektivgutexperimenten und anderen empirisch beobachtbaren Situationen wählen manche Akteure jedoch zu ihrem finanziellen Nachteil eine faire Strategie ohne Orientierung an ihren eigenen Auszahlungen. Diese Fairness mit einem geringen Grad an Zweckrationalität wird hier als Fairness ohne Eigennutz definiert. In einer Zweitauswertung einer Studie zur Gesundheitsreform wird aufgezeigt, dass ein Wechsel von der einkommensbezogenen Finanzierung der gesetzlichen Krankenversicherung in Deutschland (bisheriges solidarisches System) zu pauschalierten Beiträgen (Gesundheitsprämien bzw. Kopfpauschalen) sowohl von Verlierern als auch von Gewinnern abgelehnt wird aufgrund von Fairness ohne Eigennutz. Es wird empfohlen, die Einspeisung von Fairness ohne Eigennutz in ökonomische Modelle zu diskutieren und die Grundlagen des Paradigmenwechsels von der älteren Wohlfahrtsökonomie zur Neuen Politischen Ökonomie empirisch zu prüfen.

Abstract

Fairness for reasons of self-interest (standard economic behavioural theory) and the orientation toward fairness of payouts received (experimental economics) are seen in the economic literature as causes of co-operative behaviour. However, in experiments concerning collective goods, as well as in other empirically observable situations, some agents choose a fair strategy even if it is to their own financial detriment, independent of their own payouts. This fairness with a low level of ends rationality is defined here as fairness without regard to self-interest. In a re-analysis of a study on health reform, it is demonstrated that a shift from income-related financing of compulsory health insurance in Germany (hitherto a solidarity system) to flat-rate

1 Postprint: Biniossek, C., Betz, D., Fetchenhauer, D., & Lünen, M. (2007). Fairness ohne Eigennutz. Handlungsannahmen in der ökonomischen Theorie und Auswirkungen auf Reformoptionen im Gesundheitswesen. Sozialer Fortschritt, 56(9-10), 252-258.

contributions (uniform *per capita* health insurance premiums) is rejected by both the potential winners and losers of such a reform for reasons of fairness without regard to self-interest. It is recommended that the integration of fairness without regard to self-interest into formal economic analysis be discussed and the foundations for a paradigm shift from the older welfare economics to public choice be empirically tested.

Einführung

Aktueller Forschungsstand

In der gesundheitspolitischen Diskussion wurden in der Bundesrepublik in den Jahren bis Anfang 2006 mehrere Reformoptionen für die Einnahmen der Krankenversicherung diskutiert, insbesondere die Bürgerversicherung (Ausdehnung des Versichertenkreises und der Bemessungsgrundlage mit einkommensbezogenen Beiträgen) und Gesundheitsprämien bzw. Kopfpauschalen (Beibehaltung der privaten Krankenversicherung; pauschalierte Beiträge mit Beendigung der beitragsfreien Mitversicherung für Ehepartner) (BMGS 2003; Lungen, Gerber, Stollenwerk & Klever-Deichert 2005).²

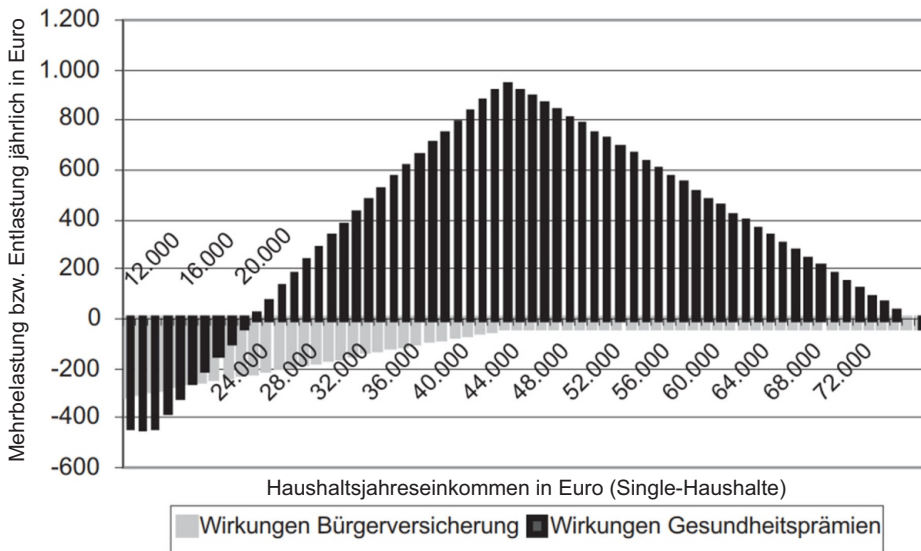
Das derzeit bestehende Finanzierungssystem ist als horizontal-genossenschaftsartige Risikogemeinschaft auf Gegenseitigkeit charakterisierbar (Schulz-Nieswandt 2002). Bisheriges Finanzierungssystem, Bürgerversicherung und Gesundheitsfonds stellen Varianten des Solidarprinzips dar und belasten Beitragszahler überwiegend nach ihrer Leistungsfähigkeit. Gesundheitsprämien bedeuten einen Wechsel vom Leistungsfähigkeitsprinzip zum Gleichheitsprinzip der Beitragszahlung (Kärcher & Kals 2004, S. 37) und werden in der ökonomischen Literatur wegen der stärkeren Trennung von Allokation und Distribution, der langfristig möglichen Entlastung der Lohnkosten der Arbeitgeber und der administrativ vereinfachten Schöpfung neuer Ressourcen zur Finanzierung des Gesundheitswesens verteidigt (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung 2004, Ziffern 307 ff.; Falk 2002, S. 57).

Unter ökonomischen Gesichtspunkten würden von jeder der genannten Reformoptionen unterschiedliche Bevölkerungsgruppen profitieren. Im Folgenden wird erörtert, welche Versicherten unter der Annahme rationalen Verhaltens für beziehungsweise gegen die jeweilige Reformoption (Gesundheitsprämien oder Bürgerversicherung) votieren müssten. Zugrunde gelegt wird dazu das standardökonomische Verhaltensmodell. Hier wird argumentiert, dass die Zustimmung oder Ablehnung ausschließlich an dem monetär ökonomischen Nutzen orientiert ist. Jedes

2 Der mittlerweile beschlossene Gesundheitsfonds als Variante der bisherigen solidarischen Finanzierung wurde erst seit 2006 diskutiert.

Individuum wird für die Finanzierungsalternative votieren, von der es sich den höchsten Gesamtnutzen verspricht.³

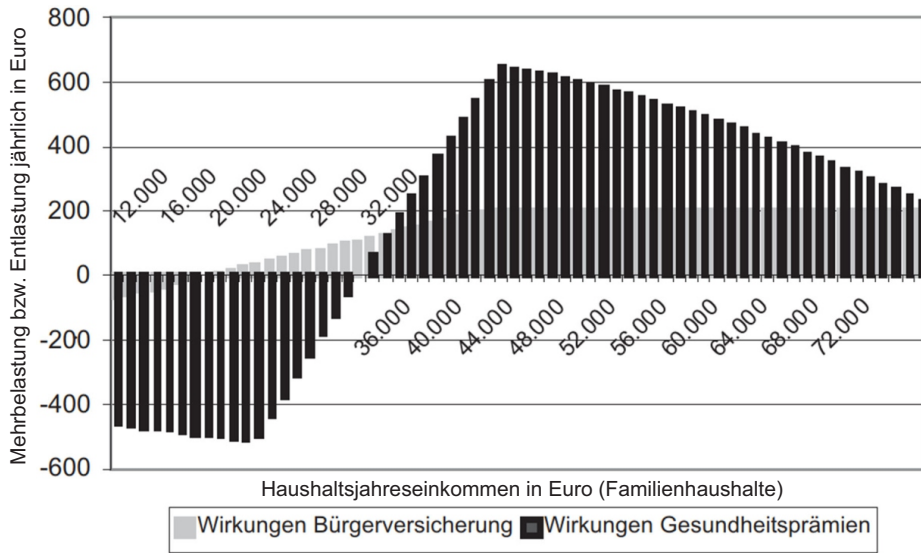
Die einkommensabhängigen individuellen monetären Nutzengewinne bzw. Nutzenverluste (relativ zum bestehenden Finanzierungssystem) werden exemplarisch anhand zweier Beispielrechnungen dargestellt. Zugrunde gelegt wurden Single-Haushalte (Grafik 1) beziehungsweise Familienhaushalte (zwei Einkommen) (Grafik 2) mit jeweils 5.000 € jährlichem Zinseinkommen.



Quelle: Eigene Berechnungen, Institut für Gesundheitsökonomie und Klinische Epidemiologie des Klinikums Köln.

Grafik 1 Wirkungen der Reformoptionen relativ zum bestehenden System (Single-Haushalte)

3 Gesamtnutzen, definiert als subjektiv erwarteter Barwert der Zahlungsströme in Abhängigkeit der erwarteten eigenen sozialen Lage.



Quelle: Eigene Berechnungen, Institut für Gesundheitsökonomie und Klinische Epidemiologie des Klinikums Köln.

Grafik 2 Wirkungen der Reformoptionen relativ zum bestehenden System (Familienhaushalte)

Die Beispielrechnung in Grafik 1 zeigt für Single-Haushalte relativ zum bestehenden System die jährlichen Entlastungen (positiver Bereich der y-Achse) oder Mehrbelastungen (negativer Bereich der y-Achse) in Abhängigkeit vom Haushaltsjahreseinkommen (x-Achse) sowohl für den Fall der Einführung von Gesundheitsprämien (schwarze senkrechte Balken) als auch der Bürgerversicherung (graue senkrechte Balken).

In Abhängigkeit vom jährlichen Bruttoeinkommen wird deutlich, dass Bezieher geringer Einkommen aus abhängiger Beschäftigung von Gesundheitsprämien belastet werden, Bezieher hoher Einkommen jedoch entlastet. Der Effekt beruht wesentlich auf der Pauschalierung der Beiträge. Die Bürgerversicherung belastet insbesondere Bezieher hoher Zinseinkünfte durch deren Verbeitragung, entlastet hingegen durch die Absenkung der Beitragssätze. Insgesamt ergibt sich bei der Bürgerversicherung eher eine Entlastung von Beziehern geringer Einkünfte und Familien ohne Zinseinkommen.

Keine der beiden Reformoptionen (Bürgerversicherung oder Gesundheitsprämien) kann eindeutig eine wesentlich größere Gruppe von Personen bei- oder entlasten.⁴ In Umfragen äußern Befragte jedoch mehrheitlich eine starke Präferenz

4 Vertreter des Gesundheitsprämienmodells (CDU) sowie Vertreter der Bürgerversicherung (SPD) behaupten jeweils, dass als monetäre Umverteilungswirkung ihres eigenen Reformvorschlags eine größere Gruppe an Personen entlastet wird.

für einkommensbezogene Beiträge. Dies geht über die Bevölkerungsanteile, welche von einkommensbezogenen Beiträgen profitieren würden, hinaus.

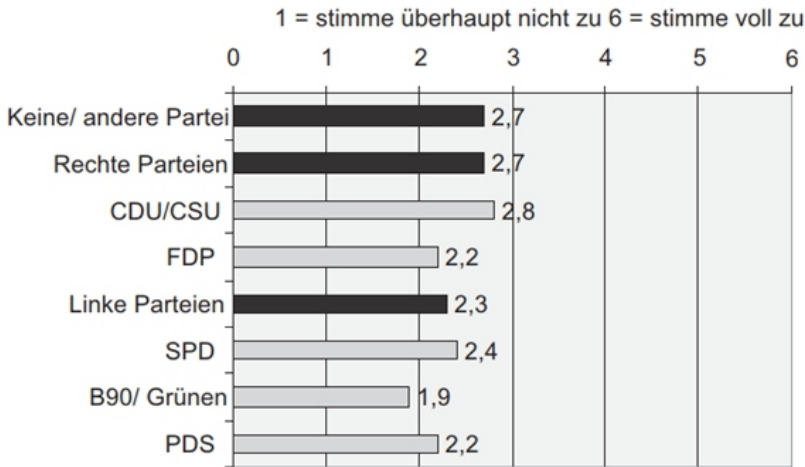
Unter der Hypothese der standardökonomischen Verhaltenstheorie, dass Personen ihre monetären Auszahlungen maximieren, sollten Gewinner und Verlierer jeder Option (Grafik 1 und 2) den für sie finanziell vorteilhaftesten Vorschlag präferieren. Konkret: Wenn man durch Gesundheitsprämien viel spart, sollte man diese Option auch wählen.

In Grafik 3 wird ersichtlich, dass pauschale Beiträge unabhängig von der Parteipräferenz stark abgelehnt werden. Auffällig sind hier die höchsten Ablehnungsraten (auf einer Skala von 1 bis 6) bei Anhängern von Bündnis90 / Grüne (1,9) und FDP (2,2), also Parteien mit überdurchschnittlich einkommensstarker Klientel,⁵ die auch den größten finanziellen Vorteil von pauschalen Beiträgen hätten. Wenn Personen gegen pauschale Beiträge votieren, von denen sie eindeutig große finanzielle Vorteile hätten, widerspricht dies dem ökonomischen Erklärungsansatz.

Die Ablehnung pauschaler Beiträge scheint zumindest nicht unmittelbar über monetären Nutzen erklärbar zu sein. In der Literatur werden verschiedene Gründe für Handlungsabweichungen gegenüber der ökonomischen Verhaltenstheorie diskutiert. Grundsätzlich lassen sich diese Anomalien in zwei Klassen unterteilen. Abweichungen zur standardökonomischen Verhaltenstheorie ergeben sich zum einen durch die Art und Weise, wie Menschen Informationen aufnehmen und verarbeiten. Hierbei können Anomalien entstehen aufgrund peripherer Informationsverarbeitungsprozesse oder mangelnder Informiertheit. Die zweite Ursache für Anomalien sind individuelle Präferenzen, beispielsweise die Präferenz für eine faire Option.⁶ Empirische Befunde der Experimentellen Wirtschaftsforschung legen die herausragende Bedeutung von Fairness bei Kollektivgutentscheidungen nahe und verweisen explizit auf eine mögliche Anwendbarkeit in der Gesundheitspolitik (Bolton & Ockenfels 2003, 1). Auch die psychologische Gerechtigkeitsforschung betont Fairness als Hauptfaktor für Abweichungen vom Eigennutzaxiom bei Einstellungen zu Reformoptionen bei der Finanzierung der Krankenversicherung (Kärcher & Kals 2004). Im Folgenden soll daher Fairness als mögliche Ursache der Ablehnung pauschaler Krankenkassenbeiträge untersucht werden.

5 Klein & Falter 2003; Der Spiegel 2004.

6 Einige Anomalien stellen Mischformen dar. Beispielsweise ist der Status Quo Bias (Präferenz für eine bereits bestehende Option, hier: etablierte Option des Bismarckschen Sozialsystems) ein Hybrid aus schematischer Informationsverarbeitung und der Präferenz für eine bereits bestehende Option. Wording (hier „Kopfpauschale“ als negativ assoziierter Begriff) stellt eine weitere Anomalie mit Hybridcharakter dar.



Quelle: Christoph / Ullrich (2006), S. 80, N (ungewichtet) = 1279.

Grafik 3 Zustimmung „Pauschale Beiträge sind besser als bisheriges System“ nach Parteipräferenz

Fairness

Während ältere wirtschaftswissenschaftliche Konzepte untersuchen, unter welchen Bedingungen faires Verhalten zu einer Maximierung des monetären Eigennutzes führen kann, fundieren neuere Konzepte der Experimentellen Wirtschaftsforschung meist auf einer Erweiterung des Rational Choice Konzeptes. Nicht-monetär äquivalente Bedürfnisse wie Fairness werden in einen weiten Nutzenbegriff integriert. Die ältere wirtschaftswissenschaftliche Fairnessforschung nimmt monetären Eigennutz⁷ als alleiniges Motiv an, definiert jede Entscheidung unabhängig von deren Grad an Fairness bzw. Kooperation als eigennutzmaximierend und führt theoretische Analysen sowie spieltheoretische Simulationen durch. Die Experimentelle Wirtschaftsforschung definiert Fairness anhand von Experimenten mit anonymen realen Auszahlungen von Bargeld. Empirische Befunde aus dem Bereich der Experimentellen Wirtschaftsforschung und der psychologischen Gerechtigkeitsforschung konnten Verhalten aufzeigen, welches für die befragte Person mit zum Teil hohen finanziellen Einbußen zugunsten anonymer Anderer verbunden ist und das mit den bisherigen Ansätzen nicht erklärbar scheint.

7 Maximierung der Befriedigung von Bedürfnissen durch Geld (als generalisiertem Verstärker).

Fairness aus dem Motiv „Eigennutz“

Fairness aus Eigennutz soll definiert werden als kooperatives Verhalten,⁸ das den Eigennutz maximiert. Eigennutzmaximierung kann hierbei unmittelbar aus der maximalen monetären Auszahlung, aber auch aus der subjektiv insgesamt erwarteten langfristigen Gesamtauszahlung bestehen. So kann beispielsweise eine Person finanzielle Nachteile aus einkommensabhängigen Beiträgen haben und diese dennoch aus Erwartung (z.B. Heirat oder Erziehungszeit) oder Unsicherheit (z.B. Zukunftsängste vor Krankheit oder Arbeitslosigkeit) bezüglich zukünftiger sozialer Lagen präferieren. Fairness kann auch als Strategie eingesetzt werden, um instrumentell an eine Belohnung zu gelangen, die größer ist als die unmittelbar realisierbare maximale Auszahlung.⁹ Die erwartete Sanktionierung (Belohnung oder Bestrafung) seitens des sozialen Umfeldes für wertekonformes bzw. sozial erwünschtes (z.B. faires) Verhalten hat direkten Einfluss auf die individuellen Nutzenkalküle. Becker (1982) legte dar, dass z.B. Altruismus (S. 320 ff.), aber auch Liebe und Ehe einen monetär äquivalenten Nutzen (soziales Einkommen) stiften können.¹⁰ Zudem kann mit einer fairen Alternative zusätzlich ein individueller Nutzen verbunden sein (Kuppelprodukte bei Kollektivgütern [Olson 1968]) und selbst im Gefangenendilemma kann durch Iteration (mehrfaches Spiel bzw. Interaktion mit gleichen Partnern) Fairness und Kooperation auch zwischen monetär eigennützigen Akteuren entstehen (Axelrod 1984).

Fairness in neueren Modellen der Experimentellen Wirtschaftsforschung

In der Experimentellen Wirtschaftsforschung wurden zu dem Motiv Fairness aus direktem monetären Eigennutz noch zwei weitere Motive zugefügt. Dies ist einerseits das Motiv nach einer gerechten Verteilung (distributive Fairness). In distribution models (Fehr & Schmidt 1999; Bolton & Ockenfels 2000) vergleichen Akteure ihre eigenen Auszahlungen relativ mit den Auszahlungen von anderen und bemessen danach Fairness. Zum anderen existiert ein Motiv nach Vergeltung einer zuvor erfahrenen positiven oder negativen Auszahlung (reciprocal kindness model) (Rabin 1993). Die Akteure sind bei beiden Konzepten interaktiv orientiert, betrachten jedoch Fairness nur bezogen auf sich selbst (self centered fairness) (Bolton & Ockenfels 2002) und vermeiden es, sich finanziell schlechter zu stellen (Güth, Kliemt & Ockenfels 2002, S. 11).

8 Entscheidung eines Individuums für eine angemessene (John Rawls) bzw. einer Person entsprechend des gemeinschaftlichen Konsenses zustehende (Jeremy Bentham) Verteilung von Gütern.

9 Z.B. das Verhalten von Politikern (Downs 1957).

10 Hierbei wird in der wirtschaftswissenschaftlichen Literatur sowohl der kardinale Nutzen (Messprobleme, Problembereich der willkürlichen absoluten aber rationalen relativen Bewertung) sowie die Rolle von Altruismus kontrovers diskutiert.

Fairness ohne Eigennutz

Als Fairness ohne Eigennutz soll im Folgenden eine Entscheidung definiert werden, bei der unter finanziellen Nutzeneinbußen keine entsprechenden Nutzengewinne erzielt werden, sondern nur eine gerechtere Verteilung der Auszahlungen erreicht wird. Fairness ohne Eigennutz liegt damit vor, wenn sich ein Individuum für kurz- und langfristige finanzielle Verluste entscheidet und dies weder durch standardökonomische Modelle noch durch neue Fairness-Modelle (distribution model, reciprocal kindness model) erklärbar ist. Fairness ohne Eigennutz stellt eine Anomalie zum ökonomischen Verhaltensansatz dar.

Empirische Untersuchungen zum Thema Fairness ohne Eigennutz

Fairness ohne Eigennutz in der Experimentellen Wirtschaftsforschung

Zunächst soll anhand einer Sekundärauswertung der Studie von Bolton und Ockenfels (2003) der Fragestellung nachgegangen werden, ob und zu welchem Grad das theoretische Konstrukt „Fairness ohne Eigennutz“ eine Strategie von Individuen in Untersuchungen der Experimentellen Wirtschaftsforschung darstellt.

Bolton und Ockenfels (2003) führten im Rahmen eines Experimentes zu Fairness bei Kollektiventscheidungen folgendes Spiel („Game II“) mit Auszahlung realen Geldes durch: Studenten ($N = 144$) wurden per Zufallsauswahl in anonyme Dreiergruppen aufgeteilt. Jede Gruppe konnte zwischen zwei Alternativen entscheiden: einer „gerechten“ Alternative A, mit der gleichen Auszahlungssumme an jedes Gruppenmitglied aber niedrigerer Gesamtauszahlung und einer „ungerechten“ Alternative B mit ungleicher Auszahlung an jedes Gruppenmitglied aber höherer Gesamtauszahlung (von Bolton und Ockenfels als effizientes Kollektivergebnis bezeichnet). Wenn die Mehrheit (zwei oder drei Personen) bei ihrer geheimen und anonymen Abstimmung für die „gerechte“ Alternative A votiert hatte, dann erhielt jeder der drei Spieler 13 DM (39 DM Gesamtauszahlung). Wenn die Mehrheit für die „ungerechte“ Alternative B gestimmt hatte, dann wurden insgesamt 45 DM in folgender Weise an die Gruppenmitglieder verteilt: Person 1 erhielt 27 DM, Person 2 erhielt 1 DM und Person 3 erhielt 17 DM.

Von 48 Personen in der Rolle von „Person 1“ votierten 29,17% für den fairen Auszahlungsplan A (im Folgenden bezeichnet als 1A = 29,17%) mit Aussicht auf 13 DM Auszahlung. Die restlichen 70,83% „Personen 1“ stimmten für die effiziente Option B (1B = 70,83%) mit Aussicht auf die höhere Auszahlung von 27 DM. Für die 48 Personen in der Rolle von „Person 2“ ergab sich 2A = 89,58% und 2B = 10,42%. Es votierten für 3A = 31,25% und für 3B = 68,75% der Versuchspersonen in der Rolle von „Person 3“.

Bolton und Ockenfels interpretierten ihre Ergebnisse (Tabelle 1) folgendermaßen: Individuen wählen gemäß ökonomischem Verhaltensmodell als dominante

Strategie Eigennutz, (1B = 70,83%, 2A = 89,58% und 3B = 68,75%), aber in Widerspruch zum ökonomischen Verhaltensmodell als subdominante Strategie Fairness (1A = 29,17% sowie 3A = 31,25%) und als weitere subdominante Strategie Effizienz (der Gruppenauszahlung) (2B = 10,42%).

Tabelle 1 Spieltheoretisches Kollektivgut-Experiment

	Option A (fair): (Gesamtauszahlung DM 39) Person 1: DM 13 Person 2: DM 13 Person 3: DM 13	Option A (effizient): (Gesamtauszahlung DM 43) Person 1: DM 27 Person 2: DM 1 Person 3: DM 17
Person 1	14 (29,17%)	34 (70,83%)
Person 2	43 (89,58%)	5 (10,42%)
Person 3	15 (31,25%)	33 (68,75%)

Quelle: Eigene Sekundärauswertung Datensatz Bolton und Ockenfels (2003), (N = 144)

Eine „Person 1“ wird aber nur dann Option A wählen, wenn Fairness auf Individual-ebene ihre dominante Strategie ist. Die Hypothese von *Bolton / Ockenfels*, dass Fairness eine subdominante Strategie darstellt, ist daher nur auf Aggregatebene nachvollziehbar. Folgt man der Argumentation von *Bolton / Ockenfels* weiter, dann müssten viel weniger „Personen 1“ die faire Option wählen als „Personen 3“, was empirisch nicht der Fall ist. Für „Person 1“ widerspricht es ihrer dominanten Strategie Eigennutz, auf 14 DM zu verzichten, um „Person 2“ einen Gewinn von 12 DM zu verschaffen, es widerspricht ihrer Strategie Effizienz, den Gruppengewinn zu minimieren und zudem wird „Person 3“ ein Leid in Höhe von 4 DM zugefügt. 1A bedeutet einen hohen individuellen Schaden für „Person 1“, einen Schaden für die Gesamtgruppe, einen Schaden für „Person 3“ und nur für „Person 2“ einen positiven Nutzen. Demgegenüber kann „Person 3“ für 4 DM Eigennutzverlust „Person 2“ einen Fairnessgewinn von 12 DM verschaffen und zudem eine unfaire Begünstigung von „Person 1“ verhindern. Die Wahl von 1A kann daher weder von der standardökonomischen Theorie (nutzenmaximales Konsumieren oder Spenden der 14 DM) noch von neueren Fairness-Modellen (zusätzliche Berücksichtigung von relativen Auszahlungen) plausibel erklärt werden.

Weiterhin ist nicht plausibel, warum bei 10,42% der „Personen 2“ die subdominante Strategie Effizienz (6 DM Gruppengewinn) ausreichen soll, um dafür sowohl auf 12 DM zu verzichten (Verstoß gegen die dominante Strategie Eigennutz) als auch eine völlig ungerechte Verteilung zu unterstützen (Verstoß gegen die subdominante Strategie Fairness). Da das Experiment nur Outcomes abbildet, werden die angenommenen kognitiven Prozesse aus Plausibilitätsüberlegungen abgeleitet, die bisher empirisch aber nicht validiert sind.

Fairness ohne Eigennutz in der psychologischen Gerechtigkeitsforschung

In der psychologischen Gerechtigkeitsforschung wird auf Basis empirischer Befunde diskutiert, dass Menschen nicht nur an Eigennutz orientiert sind (zur Übersicht: *Bierhoff* 2002). Das Verhältnis zwischen monetären Auszahlungen, Motiven (z.B. Fairness ohne Eigennutz) und Verhalten im Gesundheitsbereich wird beispielsweise am Phänomen der Blutspendebereitschaft diskutiert. In einer Studie von Mellström und Johannesson (2005) wurden 262 Versuchspersonen zufällig in drei Gruppen mit unterschiedlichen Auszahlungsplänen gelost. In der Gruppe ohne Bezahlung spendeten 42,70% der Personen Blut. In der Gruppe mit SEK 50 (etwa 7\$) Bezahlung spendeten 32,94% Blut. Bei wahlweiser Spende von SEK 50 an eine Wohlfahrtsorganisation oder Barauszahlung von SEK 50 spendeten 44,32% Blut.¹¹ Fast die Hälfte der Personen spendet somit kostenlos Blut (Wahl einer fairen Strategie), rein monetäre Anreize senken das Blutspendeaufkommen.¹² Miller und Ratner (1998, S. 54) konnten sogar einen noch höheren Grad an Fairness (kostenlose Blutspendebereitschaft) nachweisen. In ihrer Untersuchung wurden 56 Studenten nach ihrer Blutspendebereitschaft ohne und mit Bezahlung, sowie der vermuteten Blutspendebereitschaft ihrer Peers befragt. Von den Befragten äußerten 62,50% die Bereitschaft, kostenlos Blut zu spenden. Durchschnittlich vermuteten die Befragten indes, dass nur 32,64% ihrer Peers bereit seien, Blut zu spenden.¹³ Gemessen am tatsächlichen Blutspendeverhalten bei Mellström und Johannesson schätzen sich Befragte damit fairer ein, als sie tatsächlich sind, unterschätzen hingegen die Fairness anderer. Fetschenhauer und Dunning (2006) zeigten, dass Prosozialität ein sehr robustes Phänomen darstellt, Menschen jedoch die Eigennutzorientierung anderer überschätzen.

-
- 11 Von den 153 weiblichen (109 männlichen) Versuchspersonen spendeten 51,85% (28,57%) kostenlos, 30,00% (37,14%) bei Bezahlung und 53,06% (33,33%) bei wahlweiser Bezahlung oder Geldspende Blut. Der Grad an Fairness variiert damit nicht nur aufgrund situativer Faktoren, sondern auch aufgrund des Geschlechts.
 - 12 Dies wird beispielsweise als Verdrängung einer intrinsischen (hier: Motiv Fairness) durch extrinsische Motivation (hier: Motiv finanzielle Entlohnung) bzw. Crowding-out-Problematik diskutiert (Sundmacher 2006, S. 173).
 - 13 Die Wirkung ökonomischer (eigennutzsteigernder) Anreize auf das eigene Verhalten wird von Befragten jedoch völlig falsch eingeschätzt. In der Untersuchung von Miller und Ratner stieg bei einer Entlohnung in Höhe von 15\$ die geäußerte eigene Blutspendebereitschaft auf 73,21%; die gegenüber den Peers vermutete Blutspendebereitschaft stieg auf 62,46%. In der Untersuchung von Mellström und Johannesson war hingegen bei monetären Auszahlungen die tatsächliche Anzahl an Blutspendern um 9,76% geringer als ohne monetäre Anreize; Befragte schätzen sich (bezogen auf ökonomische Anreize) in der Umfrage von Miller und Ratner (im Gegensatz zu vielen anderen Umfragen) eigennutzorientierter ein, als sie tatsächlich sind (vgl. LaPiere [1934] zu prosozialem Verhalten trotz inverser Verhaltensabsicht).

Fairness ohne Eigennutz als Erklärung der starken Ablehnung pauschaler Krankenkassenbeiträge

Um den Einfluss von Fairness ohne Eigennutz auf das Votieren für Reformvorschläge zur Krankenkassenfinanzierung zu bestimmen, wurde der Datensatz „Untersuchung Gesundheitssystem“ von Pappi und Shikano (2004)¹⁴ einer Sekundäranalyse unterzogen und die Ergebnisse teilweise argumentativ durch Befunde weiterer Studien ergänzt.

Eigennutzhypothese und Ablehnung pauschaler Beiträge

Von den 1.212 Befragten fanden 19,1% pauschale Beiträge besser und 54,3% schlechter als das bisherige System; 24,7% gaben an, dies nicht beurteilen zu können und 1,9% machten keine Angaben.¹⁵ Es könnte argumentiert werden, dass Zustimmung und Ablehnung möglicherweise äußerst schwach ausgeprägt seien und in Zusammenhang mit dem hohen Anteil unentschiedener Bürger nicht zwangsläufig ein Verstoß gegen die Eigennutzhypothese vorliegen muss. Bei Differenzierung der Urteile auf einer 6-stufigen Skala ergab sich bei den gültigen Antworten ($N = 875$) eine durchschnittliche Bewertung von 2,45 (bei einem theoretischen Skalenmittel von 3,5).¹⁶ Hierbei beurteilten 5,4% pauschale Beiträge viel besser, 8,2% besser, 12,6% etwas besser, 12,8% etwas schlechter, 21,7% schlechter und 39,3% viel schlechter als das bisherige System. Dass nur 5,4% der Befragten pauschale Beiträge als viel besser beurteilen, stellt einen Verstoß gegen die Eigennutzhypothese dar (siehe Grafik 1 und Grafik 2).

14 Zweistufige repräsentative Zufallsstichprobe bei 1.212 telefonisch befragten Personen (809 West, 403 Ost, Repräsentativität durch Gewichtung der Ostüberquotierung).

15 Wortlaut des vorgelesenen Textes: „Denken Sie jetzt bitte einmal nur an die gesetzliche Krankenversicherung. Bisher müssen alle Arbeitnehmer bis zu einem bestimmten Einkommen einen bestimmten Prozentsatz ihres Einkommens als Krankenversicherungsbeitrag zahlen. Den gleichen Beitrag muss zusätzlich der Arbeitgeber zahlen. Es gibt jetzt verschiedene Vorschläge zur Reform der gesetzlichen Krankenversicherung: Ein Reformvorschlag besteht darin, dass alle gesetzlich Krankenversicherten außer Kindern zukünftig einen *gleichen* Beitrag, unabhängig von der Höhe ihres Einkommens, zahlen sollen. Versicherte mit niedrigem Einkommen sollen dann Zuschüsse aus Steuermitteln erhalten. Finden Sie diesen Vorschlag im Vergleich zum bisherigen System besser, schlechter oder können Sie das nicht beurteilen?“ Zur Vermeidung von Reihfolgeeffekten wurde die Hälfte der Befragten zuerst nach Bürgerversicherung, dann nach Bürgerpauschale und schließlich nach pauschalen Beiträgen befragt, die andere Hälfte in umgekehrter Reihenfolge.

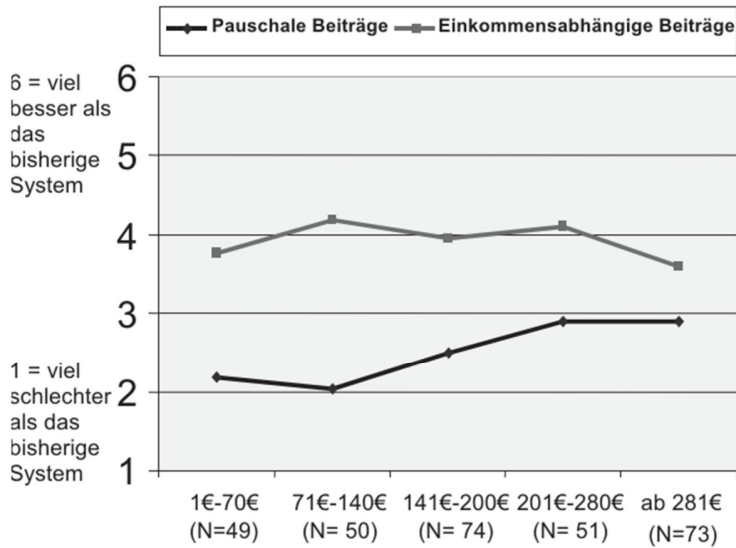
16 Befürworter und Gegner (insgesamt 73,4% der Befragten) wurden gebeten, ihre Aussage auf einer dreistufigen Skala zu differenzieren: „Und finden Sie diesen Vorschlag viel besser, besser oder etwas besser als das bisherige System?“ bzw. bei Ablehnung: „Und finden Sie diesen Vorschlag viel schlechter, schlechter oder etwas schlechter als das bisherige System?“ Hieraus wurde in der eigenen Sekundäranalyse eine 6-stufige Skala konstruiert mit 1 = pauschale Beiträge sind viel schlechter als das bisherige System, 6 = pauschale Beiträge sind viel besser als das bisherige System.

Auch in anderen Untersuchungen werden pauschale Beiträge stark negativ bewertet, unabhängig welche Begriffe verwendet werden (Gesundheitsprämien, pauschale Beiträge, Kopfpauschale), welcher Vergleichspunkt gewählt wird (bisheriges System, alternative Reformvorschläge, kein vorgegebener Vergleichspunkt) oder welche Art von Urteilen abgegeben werden sollen (Urteil zur Einführung, Gerechtigkeitsurteile). So wurden in der Studie von Kärcher und Kals (2004, S. 37) Personen befragt, wie gerecht sie das Leistungsprinzip (bisheriges System, Bürgerversicherung, Gesundheitsfonds) bzw. das Gleichheitsprinzip (Gesundheitsprämien) zur Finanzierung des Gesundheitswesens finden. Auf einer 6-stufigen Skala (1 = stimmt überhaupt nicht; 6 = trifft genau zu) bewerteten 282 gesetzlich Krankenversicherte das Leistungsprinzip durchschnittlich mit 5,01 (ziemlich gerecht). Das Gleichheitsprinzip wurde durchschnittlich mit 2,73 (relativ ungerecht) bewertet. Christoph und Ullrich (2006, S. 80) erhoben die Einstellung zu Reformvorschlägen der Krankenkassenfinanzierung auf einer 6-stufigen Skala (1 = stimme überhaupt nicht zu; 6 = stimme voll und ganz zu). Einer Ausweitung der Versicherungspflicht (Bürgerversicherung) wird mit einem Wert von durchschnittlich 4,6 zugestimmt, pauschale Beiträge werden mit einem Durchschnittswert von 2,6 abgelehnt.

Experimentelle Wirtschaftsforschung und Ablehnung pauschaler Beiträge

Die Experimentelle Wirtschaftsforschung konnte nachweisen, dass sich Individuen sowohl an monetärem Eigennutz als auch an Fairness orientieren. Die meisten Akteure verfolgen in vielen Experimenten eine monetär eigennützige Strategie, eine Minderheit verhält sich fair. Hieraus lässt sich die Hypothese ableiten, dass Verlierer von pauschalen Beiträgen gegen diese votieren, Gewinner aber großteils für und zu einem kleineren Teil gegen pauschale Beiträge stimmen. Pauschale Beiträge sollten daher von Verlierern extrem stark negativ und von Gewinnern durchschnittlich leicht positiv bewertet werden. Insgesamt würden damit in der Gesamtbevölkerung pauschale Beiträge abgelehnt. Hierbei bliebe der grundsätzliche Zusammenhang (dominante Strategie Eigennutz: je höher der monetäre Nutzen, desto größer die Zustimmung) erhalten, verschöbe sich jedoch auf ein durchschnittlich negativeres Bewertungsniveau (subdominante Strategie Fairness).

Im Datensatz wurde die Mitgliedschaft in einer gesetzlichen Krankenkasse sowie die Höhe des Krankenkassenbeitrags erhoben. Je höher der Kassenbeitrag ist, desto stärker würden Personen von der Einführung pauschaler Beiträge monetär profitieren. Aus dem Datensatz wurden daher alle kassenversicherten Beitragszahler mit gültigen Angaben zu ihren Krankenkassenbeiträgen gefiltert und in 5 etwa gleich große Beitragsklassen eingeteilt. Anschließend wurden die Präferenzen in Abhängigkeit vom eigenen Krankenkassenbeitrag für pauschale Beiträge und für einkommensabhängige Beiträge in Grafik 4 gemeinsam abgetragen.



Quelle: Eigene Sekundärauswertung Datensatz Pappi und Shikano (2004), ($N = 297$).

Grafik 4 Präferenzen und eigener Krankenkassenbeitrag

In allen Beitragsklassen befürworten die Befragten einkommensabhängige Beiträge und lehnen pauschale Beiträge ab (Grafik 4).¹⁷ Der Durchschnittswert aller Beitragsklassen zusammen beträgt für einkommensabhängige Beiträge 3,90 und für pauschale Beiträge 2,53 auf einer Skala von 1 (viel schlechter als das bisherige System) bis 6 (viel besser als das bisherige System).

Die Einschätzung von einkommensabhängigen Beiträgen bleibt über alle Einkommensklassen konstant, bei pauschalen Beiträgen besteht ein schwach positiver Zusammenhang. Konkret: Unter Kontrolle von Alter und Familienstand ergab sich eine Korrelation von .16 ($p < .01$) zwischen der Höhe der eigenen Krankenkassenbeiträge und der Zustimmung zu pauschalen Beiträgen. Selbst Profiteure von pauschalen Beiträgen bewerten diese negativer als die Bürgerversicherung. Dies widerspricht der Hypothese der Experimentellen Wirtschaftsforschung, die einen zusammengesetzten Effekt aus Fairness (durchschnittlich negative Bewertung pauschaler Beiträge) und Eigennutzmaximierung (mit den Kassenbeiträgen steigende positive Bewertung pauschaler Beiträge) prognostiziert. In Grafik 4 ergäbe sich als Prognose der Experimentellen Wirtschaftsforschung für pauschale Beiträge eine stark wachsende Zustimmung mit ansteigendem Kassenbeitrag.

Pauschale Krankenkassenbeiträge werden von einer Mehrheit der Gewinner als auch der Verlierer abgelehnt. Eigennützige Motive (monetärer Eigennutz, aber auch Fairness aus Eigennutz) scheinen nicht den erwarteten hohen Einfluss auf die starke

17 So werden in der höchsten (niedrigsten) der fünf Beitragsklassen einkommensabhängige Beiträge mit einem Durchschnittswert von 3,60 (3,76) schwach präferiert und pauschale Beiträge mit 2,89 (2,18) abgelehnt.

Ablehnung pauschaler Beiträge zu haben. Fairnesskonzepte aus der Experimentellen Wirtschaftsforschung (z.B. die oben angeführten *distribution* und *reciprocal kindness models*) sind aus zwei Gründen nicht geeignet, die empirischen Ergebnisse zu erklären: Zum einen wird auch hier Eigennutz als Hauptfaktor betrachtet, der bei der Bewertung pauschaler Beiträge jedoch von untergeordneter Bedeutung scheint. Zum anderen gehen die Fairnesskonzepte von einer Orientierung der Akteure an den eigenen Auszahlungen aus. Jedoch scheinen sich weder Gewinner, die pauschale Beiträge ablehnen, noch Verlierer, die pauschale Beiträge befürworten, an „self centered fairness“ zu orientieren.

Fairness ohne Eigennutz als Erklärung der Ablehnung pauschaler Beiträge

Aus dem Datensatz wurden vier Gruppen mit unterschiedlichen finanziellen Auszahlungen gebildet. Unverheiratete Personen, deren bisheriger Kassenbeitrag oberhalb des maximal erwarteten pauschalen Beitrages von 210 €¹⁸ liegt, profitieren in der Regel von pauschalen Beiträgen (siehe Grafik 1) und sollten daher aus monetärem Eigennutz für pauschale Beiträge votieren. Verheiratete mit einem Kassenbeitrag unter 210 € erleiden in der Regel Verluste (siehe Grafik 2) und sollten daher gegen pauschale Beiträge votieren.

Pauschale Beiträge werden in den beiden Gruppen mit einem Kassenbeitrag über 210 € mehrheitlich abgelehnt (Tabelle 2). Besonders interessant ist hier die Gruppe der nicht verheirateten GKV-Versicherten mit einem Kassenbeitrag über 210 € und daher zumeist hohen finanziellen Entlastungen bei Einführung von Gesundheitsprämien. Um ihre monetären Auszahlungen zu maximieren, müsste eine deutliche Mehrheit für pauschale Beiträge votieren. Tatsächlich votieren nur 29,9% für Gesundheitsprämien und 49,3% lehnen diese Option sogar ab.¹⁹ Personen sind damit weder an der Maximierung noch an der Fairness der eigenen Auszahlungen orientiert, sondern votieren aus Fairness ohne Eigennutz. Für die Einstellung zu Finanzierungsoptionen scheinen nicht die absolute oder relative Höhe der empfangenen Auszahlungen (optimale Zielerreichung einer maximalen Auszahlung oder einer für sich selbst fairsten Auszahlung), sondern die Fairness (Anständigkeit) des Verhaltens bzw. der Einstellung selber von entscheidender Bedeutung zu sein. Fairness, die nicht an monetärer Höhe oder Fairness der eigenen Auszahlungen orientiert ist, hier definiert als Fairness ohne Eigennutz, ist wesentlicher Faktor zur Erklärung der starken Ablehnung pauschaler Krankenkassenbeiträge in der Bevölkerung.

18 Der Bertelsmann-Gesundheitsmonitor nimmt im Herbst 2003 und im Frühjahr 2004 jeweils pauschale Beiträge von 210 € an (Böcken und Amhof 2005). Diese Annahme stellt einen Maximalbetrag dar, der oberhalb der meisten anderen Schätzungen liegt. Das Gesundheitsprämienmodell der CDU sah eine einheitliche Prämie von 180 € plus 20 € Kapitaldeckung vor.

19 Bei 19,4% unentschieden und 1,5% ohne Angaben.

Tabelle 2 Präferenzen nach Kassenbeitrag und Familienstand

	GKV-Versicherte mit Kassenbeitrag bis 210 €	GKV-Versicherte mit Kassenbeitrag über 210 €
verheiratet	19,1% (57,4%)	32,0% (55,7%)
nicht verheiratet	15,4% (68,5%)	29,9% (49,3%)

Quelle: Eigene Sekundärauswertung Datensatz Pappi und Shikano (2004). Zustimmung zur Aussage: „Gleiche Beiträge für alle gesetzlich Krankenversicherten außer Kindern sind besser (schlechter) als das bisherige System“ (N = 440).

Diskussion

In der eigenen Sekundäranalyse der Untersuchung von Bolton und Ockenfels (2003), Pappi und Shikano (2004) wie auch in anderen Studien ist neben dem Einfluss von Fairness ohne Eigennutz auch der Einfluss von (monetärem) Eigennutz evident. In der Sekundärauswertung von Bolton und Ockenfels verhielten sich (in der Rolle von „Person 1“ und „Person 3“) 30,2% der Spieler fair, aber 69,8% eigennutzmaximierend. In der Studie zur Gesundheitsreform (Pappi und Shikano) votierten 49,3% der Gewinner von Gesundheitsprämien (Singles, Einkommen über 210 €) für die faire Option und 29,9% für die eigennützige Option. Die Orientierung an Eigennutz (in Form einer steigenden Zustimmung zu pauschalen Beiträgen bei steigendem Krankenkassenbeitrag) konnte in der Regressionsanalyse bestätigt werden und kann möglicherweise erklären, weshalb im Bertelsmann Gesundheitsmonitor Herbst 2004 (Böcken und Amhof 2005) eine Kopfpauschale in Höhe von 170 € von 32% der befragten Personen als gerecht empfunden wird, eine Kopfpauschale von 109 € im Gesundheitsmonitor Frühjahr 2005 jedoch bereits von 47% als gerecht eingestuft wird.²⁰

Neben der Höhe monetärer Auszahlungen und personalen Faktoren (z.B. Geschlecht, Alter) ist der situationale Kontext ein wesentlicher Bestimmungsfaktor für den Anteil von Fairness bei Entscheidungen.²¹ Außer in den oben erwähnten

20 Im Gesundheitsmonitor Frühjahr 2005 der Bertelsmannstiftung (Böcken und Amhof 2005) wurden bei einem Beitragssatz von 109 € pauschale Beiträge von 14% als vollkommen gerecht, von 32% als überwiegend gerecht, von 22% als überwiegend ungerecht und von 18% als vollkommen ungerecht empfunden bei 10% unentschieden. [Wir danken Robert Amhof für die Bereitstellung der exakten Daten.]

21 Bei Personen, die sich freiwillig zum Höchstbetrag in der GKV versichern (statt in die kostengünstigere Privatversicherung zu wechseln und das gesparte Geld zu konsumieren oder zu spenden), steht den individuellen Mehrkosten ein marginaler Gewinn für die Solidargemeinschaft gegenüber. Demgegenüber existieren bei einer Volksabstimmung für Profiteure pauschaler Kassenbeiträge zwei mögliche Outcomes: entweder steht den individuellen Mehrkosten auch die zwangsweise Einführung einer solidarischen, fairen Krankenkassenfinanzierung gegenüber (keine Trittbrettfahrerproblematik), oder es kommt zur monetär vorteilhaften aber unfairen Option.

Ansätzen der Experimentellen Wirtschaftsforschung wird Fairness auch in neuere ökonomisch formulierte Rational Choice Konzepte der Soziologie integriert (Framing-Ansatz von Esser 2001, 2003; Zielsysteme bei Lindenberg 1993, 2006). In diesen Modellen wird darauf eingegangen, dass unterschiedliche (Fairness-)Motive je nach Situation salient sein können. Hierbei lässt sich die Ablehnung pauschaler Beiträge aus Fairness als Ergebnis einer vorgelagerten Wahrnehmungsphase und nachgelagerten Entscheidungsphase erklären. In der „Editingphase“ (Kahneman und Tversky 1979) wird eine Situation wahrgenommen und der gedankliche Rahmen (Esser 2001, 2003) bzw. das Zielsystem (Lindenberg 1993, 2006) gewählt, an dem sich das Handeln orientiert. Hierbei kann beispielsweise ein wertrationales Zielsystem aktiviert werden, das nicht an Eigennutz, sondern an einem möglichst fairen und anständigen eigenen Verhalten orientiert ist, oder auch ein zweckrationales Zielsystem, in dem eigennutzmaximierend gehandelt wird. In der nachgelagerten Bewertungs- bzw. Evaluationsphase wird nun innerhalb des gewählten Frames eine Entscheidung getroffen. Pauschale Beiträge können hierbei sowohl spontan ohne große Überlegungen (periphere Route der Informationsverarbeitung) als auch bewusst nach extensiven (wertrationalen oder zweckrationalen) Entscheidungsprozessen (zentrale Route der Informationsverarbeitung) abgelehnt werden. Die psychologische Gerechtigkeitsforschung geht davon aus, dass in verschiedenen Lebenskontexten unterschiedliche Gerechtigkeitsprinzipien Anwendung finden. Während in Arbeitskontexten eine Orientierung an Beitragsgerechtigkeit (equity Prinzip) vorherrscht, wird im Bereich Gesundheit eher das Bedürfnisprinzip vermutet (Präferenz für eine solidarische und einkommensabhängige Krankenkassenfinanzierung).

Insbesondere im Vorfeld und bei der Durchführung von Reformvorhaben (z.B. Finanzierung der Krankenversicherung, Pflege von Angehörigen, Impfung, Blutspende) ist empirisch zu prüfen, ob Fairness ein wichtiger Faktor der individuellen und kollektiven Nutzenfunktionen darstellt. Allokative ökonomische Modelle sind zu modifizieren, wenn Individuen sich gegen die eigennutzmaximale Option entscheiden und dies auf Aggregatebene zu einer relevanten systematischen Verzerrung führt. Je stärker Fairness von Bedeutung ist und je mehr Individuen sich fair verhalten, desto weniger stellt der Faktor Fairness ein unbedeutendes makroökonomisches Rauschen dar.

Fazit und Ausblick

Wir haben gezeigt, dass sich im Bereich der Beitragserhebung für gesetzliche Krankenkassen in Deutschland in Umfragen eine starke Präferenz für eine an der Leistungsfähigkeit orientierte Beitragsbemessung ergibt, die zudem auf die gesamte Wohnbevölkerung ausgedehnt würde (Bürgerversicherung). Die in der gesundheitspolitischen Diskussion alternativ vorgestellten Gesundheitsprämien fanden hingegen weniger Zustimmung. Eine Gegenüberstellung von monetärem Nutzen des

Einzelnen und seiner Präferenz zeigte, dass sich Abweichungen von ökonomischen Handlungstheorien ergaben. Trotz monetärer Vorteile votierten Befragte gegen Gesundheitsprämien. Fairness ohne Eigennutz, beschreibbar als faires Verhalten mit geringem Grad an Zweckrationalität, ist bei Umfragen zu Reformoptionen im Gesundheitswesen bei vielen Befragten die dominante Strategie für Entscheidungen und erklärt die starke Ablehnung pauschaler Krankenkassenbeiträge bei Gewinnern (einkommensstarke Personen mit bisher hohen Kassenbeiträgen) dieser Reformoption. Entscheidend dafür, dass von vielen Personen Fairness ohne Eigennutz als Strategie gewählt wird sind personale Faktoren und der situative Kontext (z.B. ein solidarisches und auf Gegenseitigkeit ausgerichtetes Gesundheitswesen). Fairness ohne Eigennutz ist insbesondere im Vorfeld und während der Durchführung von Reformvorhaben von Bedeutung. Es wird empfohlen, den Faktor Fairness in alloкатive Modelle einzuspeisen.

Empirisch wurde hier wie auch in der Experimentellen Wirtschaftsforschung falsifiziert, dass Individuen sich ausschließlich eigennutzorientiert verhalten und nie am Gemeinwohl orientiert sind. Damit entfällt das Hauptargument des Paradigmenwechsels von der normativen paretianischen Wohlfahrtsökonomie (Musgrave 1959) zur positiven Ökonomie (Downs 1957; Olson 1968; Niskanen 1971). Es ist daher empirisch zu bestimmen, unter welchen Bedingungen die Analysen der Wohlfahrtsökonomie oder die Analysen der Neuen Politischen Ökonomie Gültigkeit haben.

Die behaupteten kognitiven Prozesse wurden hier wie in der Experimentellen Wirtschaftsforschung aus Plausibilitätsüberlegungen abgeleitet, da das vorliegende Datenmaterial keine direkten Rückschlüsse auf psychische Vorgänge erlaubt. Es erscheint daher notwendig, in weiteren Untersuchungen die verschiedenen Fairnesskonzepte sowohl qualitativ als auch quantitativ zu validieren.

Literatur

- Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Becker, G. S. (1982). *Der ökonomische Ansatz zur Erklärung menschlichen Verhaltens*. Tübingen: Mohr.
- Bierhoff, H. W. (2002). *Prosocial Behavior*. Hove: Psychology Press.
- BMGS (Bundesministerium für Gesundheit und Soziale Sicherung) (2003). *Nachhaltigkeit in der Finanzierung der sozialen Sicherungssysteme*. Bonn: Bericht der Kommission. Bonn; 2003.
- Böcken, J. & Amhof, R. (2005). Finanzierungsoptionen des Gesundheitswesens aus Bevölkerungssicht. In J. Böcken, B. Braun, M. Schnee & R. Amhof (Hrsg.), *Gesundheitsmonitor 2005. Die ambulante Versorgung aus Sicht von Bevölkerung und Ärzteschaft* (S. 120-136). Gütersloh: Verlag Bertelsmann Stiftung.
- Bolton, G. E. & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity and Competition. *American Economic Review*, 90(1), S. 166-193.

- Bolton, G. E. & Ockenfels, A. (2002). Self-centered Fairness in Games with More than Two Players (#42-2002). Jena: Max Planck Institut.
- Bolton, G. E. & Ockenfels, A. (2003). The Behavioral Tradeoff between Efficiency and Equity when a Majority Rules (#12-2003). Jena: Max Planck Institut.
- Christoph, B. & Ullrich, C. G. (2006). Die GKV in den Augen der Bürger: Wahrnehmung des Lastenausgleichs und Bewertungen von Reformoptionen. *Sozialer Fortschritt*, 4 / 2006, S. 75 - 83.
- Der Spiegel (2004). Grüne lösen FDP als Partei der Besserverdiener ab. *Der Spiegel*, 34 / 2004.
- Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.
- Esser, H. (2001). *Soziologie. Spezielle Grundlagen*. Band 6: Sinn und Kultur. Frankfurt am Main: Campus.
- Esser, H. (2003). Der Sinn der Modelle. Antwort auf Götz Rohwer. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 55, S. 359-368.
- Falk, A. (2002). Fairness contra Eigennutz. Konsequenzen für die ökonomische Politikberatung, In E. Fehr & G. Schwarz (Hrsg.), *Psychologische Grundlagen der Ökonomie* (S. 55-59). Zürich: NZZ-Verlag.
- Fehr, E. & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 3, S. 769-868,
- Fetchenhauer, D. & Dunning, D. (2006). Perceptions of Prosociality and Solidarity in Self and Others. In D. Fetchenhauer, A. Flache, A. P. Buunk & S. Lindenberg (Hrsg.), *Solidarity and Prosocial Behavior. An Integration of Sociological and Psychological Perspectives* (S. 61 -74). New York: Springer.
- Güth, W., Kliemt, H. & Ockenfels, A. (2002). Fairness versus Efficiency - An Experimental Study of (Mutual) Gift Giving (#40-2002). Jena: Max Planck Institut.
- Kärcher, J. & Kals, E. (2004). Gesundheitsversorgung als Konfliktfeld: Lösungsbeiträge der Gerechtigkeitspsychologie. *Wirtschaftspsychologie aktuell*, 4, S. 35-39.
- Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47, S. 263-291.
- Klein, M. & Falter, J. W. (2003). *Der lange Weg der Grünen. Eine Partei zwischen Protest und Regierung*. München: C.H. Beck.
- LaPiere, R. T. (1934). Attitudes vs. Actions. *Social Forces*, 13, 2, S. 230-237.
- Lindenberg, S. (1993). Framing, empirical evidence, and applications. In P. Herder-Dorneich, K-E. Shen & D. Schmidchen (Hrsg.), *Jahrbuch für Neue Politische Ökonomie*. 12. Band. Neue Politische Ökonomie von Normen und Institutionen (S. 11 -38). Tübingen: Mohr.
- Lindenberg, S. (2006). Prosocial Behavior, Solidarity, and Framing Processes. In D. Fetchenhauer, A. Flache, A. P. Buunk & S. Lindenberg (Hrsg.), *Solidarity and Prosocial Behavior. An Integration of Sociological and Psychological Perspectives* (S. 23- 44). New York: Springer.
- Lüngen, M., Gerber, A., Stollenwerk, B. & Klever-Deichert, G. (2005). Die Bürgerversicherung in der Krankenversicherung. *Wirtschaftspolitische Blätter*, 4/2005, S. 453-464.

- Mellström, C. & Johannesson, M (2005). Crowding Out in Blood Donation: Was Titmuss Right? (Working Papers in Economics #180) Göteborg: Göteborg University.
- Miller, D. T. & Ratner, R. K. (1998). The Disparity Between the Actual and Assumed Power of Self-Interest. *Journal of Personality and Social Psychology*, 74 (1), S. 53-62.
- Musgrave, R. A. (1959). *The Theory of Public Finance. A Study in Public Economy*. New York: McGraw-Hill.
- Niskanen, W. A. (1971). *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton
- Olson, M. (1968). *Die Logik des kollektiven Handelns*. Tübingen: Mohr.
- Pappi, F. U. & Shikano, S. (2004). Untersuchung Gesundheitssystem (Studiennummer 4330). Köln: Zentralarchiv für empirische Sozialforschung.
- Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review*, 83, S. 1281-1302.
- Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2004). Staatsfinanzen konsolidieren - Steuersystem reformieren. Jahresgutachten 2003 / 2004. Stuttgart: Metzler-Poeschel.
- Schulz-Nieswandt, F. (2002). Zur Genossenschaftsartigkeit der Gesetzlichen Krankenversicherung. Moralökonomische, kulturanthropologische und tiefenpsychologische Aspekte einer Analyse des Gabemechanismus der Sozialversicherung im morphologischen Vergleich zur Sozialhilfe. Regensburg: Eurotrans.
- Sundmacher, T. (2006). Fördern und Fordern auch in der Gesetzlichen Krankenversicherung? Eine Untersuchung auf der Grundlage unterschiedlicher ökonomischer Verhaltenstheorien. *Sozialer Fortschritt*, 7/2006, S. 168-177.

7 The Squeaky Wheel Gets the Grease – Reference Point Variations in Game Theoretic Experimental Dictator Games¹

Dirk Betz^{*#§} (corresponding), Marcel Oestreich[§] and Claudia Biniossek^{*§}

^{*} GESIS Leibniz-Institute for the Social Sciences, Unter Sachsenhausen 6-8
50667 Cologne, Germany

[#] Cologne Center for Comparative Politics (CCCP), Faculty of Management, Economics
and Social Sciences, IBW-Gebäude, Herbert-Lewin-Straße 2, 50931 Cologne, Germany

[§] Otto von Guericke University Magdeburg (OVGU), Faculty of Economics and
Management (FWW), Chair of Economic Policy, Gebäude 22, Universitätsplatz 2, 39106
Magdeburg, Germany

[§] Department of Economics, Brock University, Niagara Region, 1812 Sir Isaac Brock Way,
St. Catharines, ON, L2S 3A1 Canada

We thank the GfEW Heinz Sauermann Förderpreis and the German National Fund (DFG) for their financial support. Oestreich acknowledges financial support from the Social Sciences and Humanities Research Council of Canada under Insight Grant 435-2017-463. We also wish to thank the elfe laboratory team at the University Duisburg-Essen for helping conduct the experiments, Jeannette Brosig-Koch for her research assistance, and the participants of the Venice Seminar 2019 for very valuable remarks on this paper particularly Hartmut Esser and Andreas Diekmann.

JEL Classification: C7, C71, C91

Keywords: dictator game, laboratory experiments, model of frame selection (MFS), experimental game theory, rational choice theory (RCT)

1 Betz, D. Oestreich, M., & Biniossek, C. (2020). The Squeaky Wheel Gets the Grease – Reference Point Variations in Game Theoretic Experimental Dictator Games. Manuskript eingereicht zur Publikation bei der Fachzeitschrift *Theory and Decision*

Abstract

Findings on the experimental dictator game (EDG) report inconsistent and unstable transfer rates on both the individual and the collective level for objectively irrelevant design variations. In three variants of the EDG ($N = 72$), we test the explanatory power of the underlying decision-making mechanisms of outcome maximization in rational choice theory. We ask whether the dictator's decision passively clings to the objective stimulus or whether subjective perception and active individual cognitive processing determine the decision. We observe that larger suggested allocations lead to slightly higher transfers by dictators - in line with the saying: "The squeaky wheel gets the grease". This relationship holds for suggestions made by the recipient but also for suggestions made by the computer. However, when suggestions are made by recipients average transfers are between 42 % and 46 % larger.

Introduction

The experimental dictator game (EDG) has one of the simplest possible strategic experimental designs. Typically, in a laboratory setting, one participant (dictator) has to determine if and how much of her endowment she wants to transfer to another anonymous participant (recipient). The recipient cannot react but has to accept the decision made by the dictator. Hence no punishment is possible. This experimental design, thus very simple, has opened an ongoing academic debate, including a large number of experiments, scientific papers, and discussions.

In the beginning, when Kahneman, Knetsch, and Thaler (1986) started the dictator game paradigm, which was then altered by Forsythe, Horowitz, Savin, and Sefton (1994), research was about the fact that dictators transferred some of their endowment to the recipients. These results contradict game-theoretic predictions based on the assumptions of narrow versions of rational choice theory (RCT) of maximizing the expected monetary outcome as strictly selfish behavior. On an aggregate level, it turns out that dictators transfer, on average, 28.35 % (mean) of their endowment to the recipient (Engel, 2011).

As EDG results became more robust, specifically the observation that dictators act against game-theoretic predictions, theories of social preferences kept the basic principle of RCT that dictators choose the alternative which maximizes the expected utility but do argue that dictators often are not only interested in their personal benefits (selfish preferences) but also in the earnings of the recipient (social preferences). The personality and the degree of social preferences such as altruism, inequality aversion or reciprocity, decide how much dictators transfer to a recipient. For example, keeping 70 % and giving 30 % might maximize the individual combined selfish and social preferences of a dictator better than taking all the funds or choosing the equal split. Later theories integrated evidence from experimental market games suggesting that social preferences barely occur in market situations, by modeling social preferences dependent upon the type of (game) situation within rational

choice theoretic models (Bolton & Ockenfels, 2000; Cox, Friedman, & Sadiraj, 2008; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Fehr & Schmidt, 1999). These more recent variants of RCT still follow the computer analogy in modeling human decision-making mechanisms and analyze the EDG based on the axioms of the induced value theory (Smith, 1976): In a given EDG situation (independent variable) the subjective perception and framing of the EDG are simultaneously induced (independent variable induced by the experimenter) and the given personality of the dictator (independent variable) defines how much she suffers from inequality in this specific situation and which split of the endowment she prefers (dependent variable).

Even if the above-mentioned theories on the theoretical basis of social norms and social preferences such as altruism or inequality aversion could explain the fact that a dictator gives money to her anonymous recipient (instead of offering it to some poor person she knows or encounters in the street), these theories cannot explain further (new) experimental data pointing out the EDG's high design sensitivity. In similar EDG situations with given personalities of players, transfers should not vary across time and objectively irrelevant design variations. Brosig-Koch, Riechmann, and Weimann (2007) find that the dictator's intrapersonal behavior in the EDG is neither consistent nor stable over time. Specifically, the behavior is only stable for selfish behavior and not for prosocial behavior. Contrary to the predictions of the above-mentioned RCT inspired theories, there is a body of empirical evidence that mere design changes do change the EDG outcomes. Examples are role uncertainty during the actual decision-making (Walkowitz, 2018), or framing effects regarding *give* or *take* options (Bergh & Wichardt, 2018), thus framing which does not change the situation objectively. Krupka and Weber (2013) point out the role of social norms in EDG decisions in seemingly identical situations. In Yamamori, Kato, Kawagoe, and Matsui (2008), recipients could tell dictators anonymously the minimum share they would be willing to accept or, as an additional option, decide not to tell their minimal acceptable offers ("not to tell" option). As the recipient's choice still is not binding for the dictator, the authors report transfer rates depending on the demands of the recipients with a peek at the equal split. Compared to standard EDG, transfers were higher around the demand for the equal split, but lower in most other cases. Furthermore, the experimental design, which frames a dictator's decision as *give* to the recipient or *take* from the recipient, generates a change in results (Zhang & Ortmann, 2014; Engel, 2011; Bardsley, 2008; List, 2007)². However, contrary to these results, Dreber, Ellingsen, Johannesson, and Rand (2013) find no significant effect in their *give* and *take* studies. Framing the design, such as *giving to a charity*, does increase the transfers to the recipients (Grossman & Eckel, 2015; Bardsley, 2008). Cartwright (2016) points out that this contradictory evidence could result from a blurred definition of the frames. However, it could be additionally argued that even with a large

2 Though, Goerg, Rand, and Walkowitz (2019) find that framing has an impact on prisoner's dilemma decisions but not in dictator games.

and growing body of empirical studies on the EDG, theoretical contributions are sparse (Gerlach & Jaeger, 2016).

Still, it is remarkable that the EDG, in its simplicity, appears too complex to be explained on a theoretic level. In the literature, the hitherto unsuccessful search for an explanation of the confusing results remains limited to situational (e.g., framing) or personal factors (e.g., social preferences, personality traits), because argumentations do not question the following basic assumptions: (1) paradigm: induction of the payout structure as utility structure, (2) existence of (social) preferences, (3) selection mechanism “optimization” and (4) (temporal) causality between decision and behavior.

Thus, taking the above-cited study of Yamamori et al. (2008) as a starting point, the present work aims to formalize the boundaries of dictator *giving* and questions the underlying decision mechanism. More precisely, our research question is whether players in EDG maximize their utility function or if they activate matching mental models (social perception, framing, situational personality traits, principles of justice, attitudes, decisions, behavioral scripts). An overarching research question is whether and in what way players react to stimuli with no influence on utility maximization calculations, but which might have a significant influence on the subjective perception of the situation.

To answer this question, we first discuss the differences between the underlying basic selection principles. Then, we derive general and specific hypotheses for our game design, present our method and report and discuss our results.

Theory: Utility maximizing versus matching

In most behavioral theories, the commonly assumed underlying decision mechanism is outcome maximization. Thus, this mechanism is the basis for rational choice theory (RCT). For all theories based on RCT’s optimization mechanism, the basic principle is maximizing the product of beliefs and preferences for future consequences. Following the standard economic model, economic utility theory (von Neumann & Morgenstern, 1944), Rational Choice Theory (RCT1, RCT2, RCT3) (Edwards, 1954; Esser, 2018; 2019), this means $e \cdot U$, where e denotes the agent’s expectation for a given situation typically measured in probabilities. U denotes the agent’s utility in a given situation typically measured in monetary units.

Most experimental economists derive the explanation of the motivation (mental process) from the performed behavioral action of the players. Thus, the definition of a situation and the personality (e.g., altruistic personality) are considered as independent variables. The choice between alternatives (e.g., equal split) is considered as dependent variable. The axiomatic definition of the game situation and the axiomatic definition of an intersituative persistent personality as independent variables leave no space for any explanation which would cover different choices of the same person in the same EDG situation if only objectively irrelevant parameters are changed.

Due to the inexplicability of a huge quantity of experimental results, in the empirically oriented economic and social sciences it is discussed whether narrow versions of rational choice theory (RCT) need to be extended, such as in (Cumulative) Prospect Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), if the underlying mechanism of economic utility theory needs to be changed or if it needs to be abandoned. In the EDG, the factor risk is excluded as dictators do have 100 % control over their decision and the respective outcome. Within RCT oriented explanations, if a dictator chooses to give 10 %, she is mainly selfishly motivated, but also satisfies her (relatively weak) social preferences such as altruism or a preference for giving. This interpretation is based on the axioms of the induced value theory (Smith, 1976) and plausible reasoning. Example: As social preferences are defined by the performed action, this would lead to the interpretation that a mother who gives 10 % of her cake to her child and takes the other 90 % for herself is categorized as having at least some social preferences.

But within the framework of the model of frame selection (MFS) (Esser, 1996; 2001a; 2001b; 2002a; 2002b; 2003; 2010a; 2010b; 2018; 2019; Kroneberg, 2005; 2011; 2014; Esser & Kroneberg, 2015), the logic of the situation is the dependent variable and therefore it is not possible to deduct mental processes from the mere behavioral outcome. The above-mentioned mother might be motivated by weak social preferences such as nasty (Abbink & Sadrieh, 2009) or sadistic motives of humiliating her child or on the contrary by strong social preferences (to prevent the child from eating too much sugar). The insight here is that a participant's social preferences cannot necessarily be derived from observed behavior. Vernon Smith (2010) raised the question if the game intended by the experimenter is really induced in the heads of the players.

The *match* (m) as the basic mechanism of the (MFS) is based on the decoding of the subjectively relevant stimuli of a situation. The stimuli (cues) activate situation-specific mental models such as the perception and framing of a situation, from which the actor selects the matching (most appropriate) mental model. This subjective definition of the (game) situation then triggers the appropriate behavioral scripts. Mental models, framing, emotions, goals, preferences, habits, and behavioral scripts are all encoded by biological predisposition and socialization (Esser, 2017). "The set of mental models for different situations makes up the (multiple) identity of actors." (Esser, 2017, p. 8). This means that the *definition of the situation* is not induced by the experimenter but chosen by the individual player. Furthermore, the personality is less a fixed set of strategies but rather the choice of adequate situational roles, values, and reaction patterns. Therefore, minor cues in situational variations can lead to major differences in the dependent variables definition of a situation and the situational type of personality and, thus, ultimately to another behavior. This predicts different possible reactions of the same individual in marginally different situations as well as the change of reactions over time caused by the variation of minor situational cues (stimuli), which are irrelevant in RCT. The MFS defines this mechanism as *reframing*.

The MFS's basic selection principle is the *match*, consisting of situational stimuli (bottom-up processes) and the appropriate matching mental models (top-down pro-

cesses). The core factors for the *match* (m) of a mental model respectively its resilience are (1) the accessibility (a) of the mental model, (2) its salience in a specific situation with presence of a significant symbol (e), and (3) the absence of disturbing stimuli (u) (Esser, 2001b, p. 270). These factors are multiplied in the MFS, giving $m = a * e * u$, with $m = 1$ as the so-called *perfect match*. In the MFS, once there is a *match* of a mental model with the situation, the corresponding behavioral scripts which are available and fit to the mental model are chosen on the peripheral (as-mode) versus central (rc-mode) route of information processing. This means that even if a dictator decides to give nothing, the corresponding behavior (pulling the trigger of *giving nothing*) can only be performed if this is an existing script in this social situation and the activated situational personality. The model's core assumptions imply for the EDG that (1) a dictator chooses her subjective *definition of the situation* (2) activates a situational specific personality, and (3) then selects the appropriate principle of action within the set of given mental models and scripts. Methodologically this means that perception, personality, and the principle of action are the dependent variables.

We introduce the MFS core component *match* as an explanatory framework for the heterogeneous data derived from different EDG design variants³. We use Coleman's bathtub as our explanatory schema to illustrate the differences between the maximizing of the expected outcome (RCT) and the matching of top-down mental models with bottom-up stimuli (MFS). Figure 1 highlights the basic mechanism of the MFS for dictator games by using Coleman's bathtub as a schematic diagram:

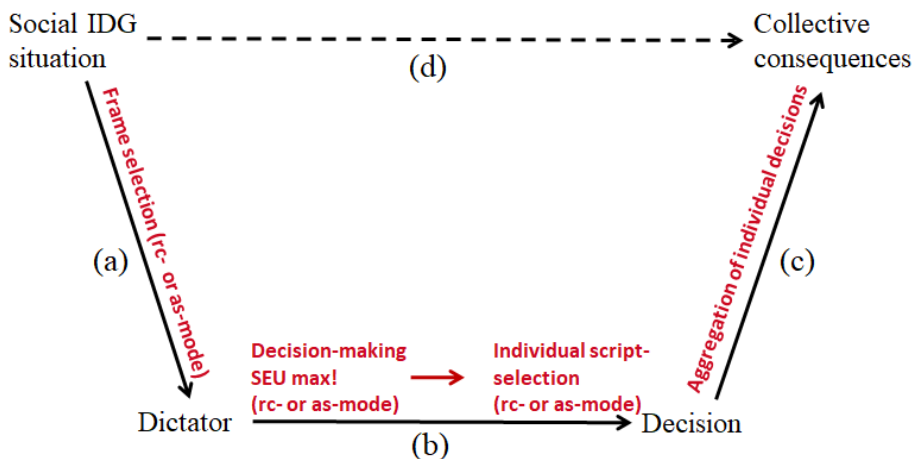


Figure 1 Illustration of Coleman's bathtub (see Esser, 1999, p. 98)

3 Nevertheless, all models including the MFS have to withstand the evidence from neuroscience, which states the inverse causality of decision making and action (Frith & Haggard, 2018; Haggard & Eimer, 1999; Kornhuber & Deecke, 1965; Libet, Gleason, Wright, & Pearl, 1983; Nann, Cohen, Deecke, & Soekadar, 2019). Thus, taking the matching principle of the MFS seriously, the situational personality and possible scripts immediately match, then the matching behavior is performed, and at least the perceived own behavior matches with the perception of one's own urge to act.

Figure 1 illustrates the social situation of a typical EDG in which every player (a) defines the subjective interpretation of the EDG situation. If there is need for cognition, the evaluation of the appropriate interpretation of the situation might be an extensive process. A dictator might for example ask herself whether the instructions are true, whether the recipient will suffer, whether the situation is about competition or about game-theoretic knowledge, or what the experimenter or her recipient expects her to do. Esser (1996; 2001b; 2002a; 2002b; 2003; Rohwer, 2003) formulates decisive factors that lead to the *match* of a mental model, respectively, to the *reframing*.

We hypothesize that the heterogeneous EDG data can be explained more efficiently within the MFS framework filled with specific principles of action than within RCT. For example, the finding that the behavior in an EDG with charity framing is less sensitive to the *give* and *take* framing than in a normal EDG is explained by Grossman and Eckel (2015) as caused by social context. Grossman and Eckel argue that a standard EDG is artificial, whereas giving to a charity is part of everyday life. This means that the frame giving to a charity provides a framework as an orientation on how to act. However, their statement concerning the role of social norms describes the phenomenon ex-post based on plausible reasoning and does not explain the mechanisms ex-ante. Concepts of social cognition (psychology) and principles of justice (Biniossek & Betz, 2018) filled in the framework of the MFS would have hypothesized ex-ante that behavior in a charity frame versus a neutral frame or in a *give* frame versus a *take* frame is structurally different due to different mental models that are activated in the experimental situation. Therefore both, less and also more sensitivity to the *give* and *take* framing are possible. The effect depends on the specific experimental design. Thus, strong habits in a situation would not reduce but increase the design sensitivity if the interaction effect with a *take* versus *give* framing activates more heterogeneous decision patterns on an individual and aggregate level.

Hypotheses

To test hypothesis derived from psychological concepts (e.g., cognitive processes of perception, categorization or theories of justice) implemented in the MFS framework and its core mechanism *match*, this paper presents the results of an innovative EDG with structural similarity to Yamamori et al. (2008), but without the authors' interpretation as communication effect and without the additional "not to tell" option. As we have shown above, the subjective definition (interpretation) of a situation according to the MFS is not induced by the given payout scheme nor the wording and framing in the instructions, but it depends on the player's subjective *match* of the situation with her mental models. We vary the basic setup of a dictator game in two different ways, which are nevertheless objectively equal. According to RCT, there should not be any behavioral differences between the outcomes of the standard EDG and our dictator game variations. The normative prevision of RCT is that in an EDG

the dictator would give nothing to the recipient as the rational choice. Though, as illustrated above, the mean transfer rate from dictators to their recipients is, on average, 28.35 % of the endowment (Engel, 2011) but can vary depending on design variants and on different framings of the experimental situation.

The control of decision behavior is based on the complex interaction of the cognitive system's many different processes. Situational characteristics, mental models, behavioral reactions, and their consequences are integrated into one episode by the cognitive system as modeled within the BRAC framework (Frings et al., 2020). The MFS is an explanatory framework that integrates intentional (top-down) and automatic (bottom-up) influences on the integration and retrieval of mental models. The definition of mental elements is open to cover hypothetical constructs, such as attributes, decisions and complex episodes. The MFS does not predict the specific content of an attribute or mental model. There is a variety of medium-range approaches that specify this relationship between stimuli, attributes, and activated mental schemes. Nevertheless, the MFS does explain the functional principles of how the stimuli-appropriate framing, goal systems, and decisions are retrieved. Subjectively, there might be different definitions of the game situation. In each way of perceiving and framing a situation, there exists a set of adequate principles of justice (see Cook & Hegtvedt, 1983; Deutsch, 1975; Rawls, 1971), norms, and behavioral scripts. Initially, there is a resource that "falls from the skies into the hands of the dictator". A dictator can subjectively define this situation as a win in a lottery or finding coins on the pavement, meaning that the dictator possesses the property rights of that resource. The corresponding principle of justice might be self-interest, which also covers the principle of *equity* as the finder (dictator) has luckily "earned" his treasure and, therefore, takes all the money. Another possible definition of the situation could be Christian charity, resulting in giving a small amount of money to the recipient as a donation. The dictator could also follow the idea of the *need principle of distributive justice* and choose an external option to give some or all of the money to a person in need or to a charity. One further alternative subjective definition of the situation is that the dictator frames the game situation as a common earning of her group of two with herself in the power of decision-making. Within this subjective mental model, the corresponding principles of justice might be equality (equal split), self-interest, or competition (taking more than the recipient).

To examine if these concepts implemented into the theoretic framework of the MFS can provide a better explanation of EDG results than RCT and in order to learn more about a dictator's decision reasons, we create two variants of the standard EDG that do not change the decision situation objectively but only subjectively. After the dictator's decision, we additionally ask for her decision motives.

Hypothesis 1 (Social situation induced by IDG)

We call the first variant *Inverse Dictator Game* (IDG) and design this treatment within an experimental setting similar to the one of Yamamori et al. (2008). The central idea of the IDG is that the recipient can propose the split of the endowment she prefers in strategy mode (Selten, 1967; Mitzkewitz, & Nagel, 1993). After receiving the recipient's proposition, the dictator decides which final decision she takes in response to the recipient's proposition. Thus, we create a minimal social situation, which (a) suggests that the money is a common resource with the dictator only in charge of allocating the resource between both players and equality (equal split) as the appropriate principle of justice. This kind of veto power (b) implies an ethical responsibility for the dictator to synchronize with the recipient's proposition. In literature, giving the dictator more right[s] to choose (Heinrich, Riechmann, & Weimann, 2009) empirically leads to higher transfer rates as a social situation is induced. Depending on the recipient's proposals, Yamamori et al. (2008) also report increasing transfer rates (see Figure 2) in their variation of a split strategy mode EDG with minimally acceptable offers and a no-communication option ("not to tell"). However, the authors explain the effect by minimal communication (voice) and not by the induced social situation. Thus, following the ideas of the MFS, we assume that in the IDG, the principle of justice *equality* is the mental model strongly supported by the situational stimuli (social situation of a common pool). In the IDG there should be a significantly higher number of individuals choosing the social situation and the *equality* principle as the appropriate *definition of the situation* than in standard EDG where property rights and self-interest are dominant. Even though objectively, the situation does not change. The behavioral reason *fairness* should be more important in IDG than in less social situations like the standard EDG.

Hypothesis H1: In the *Inverse Dictator Game* (IDG), higher transfers are realized compared to standard EDG with fairness as a prominent decision reason.

Hypothesis 2 (Non-social situation with status quo effect in RPV)

In the second variant of the EDG, called *Reference Point Variation* (RPV), the distribution is not suggested by the recipient but by the computer. Here, the dictator finds one part of the money in her account and the other part of the money in the recipient's account. As taking money from the account of another person should make it mentally more difficult for the dictator to frame the situation as having full property rights over the money in both, her account and the account of the recipient, transfer rates should also be higher than in the baseline EDG, but lower than in the social situation of the IDG. The cue that a part of the money is in the account of the recipient supports the mental model that the dictator does not have the property rights over the full amount of money. This social cognition argument within the framework of the MFS is also supported by the argumentation of the status quo bias (Samuelson & Zeckhauser, 1988; Kahneman, Knetsch, & Thaler, 1991) which can be interpreted as a mental model. As the distribution given by the computer is a refer-

ence point for the dictator, which should not activate negative emotions, this absence of disturbing stimuli strongly supports the mental model of the recipient's property rights. The decision reasons for dictators in the RPV should thus be more centered on self-interest because of the more neutral setting.

Hypothesis H2a: We hypothesize that when the dictator in both RPV treatments, RPV1-9 and RPV5-5, has to take money from the recipient, she will transfer more money compared to benchmark EDG.

As the computer makes the initial allocation in RPV, no social situation is induced. Therefore a less social situation compared to IDG leads to less activation of a common pool association, less activation of the *equality* principle, and less social bonding between dictator and recipient. All three effects lead to lower transfers in RPV5-5 than in IDG5-5 and in RPV1-9 versus IDG1-9.

Hypothesis H2b: We hypothesize that if the initial distribution is made by the computer, the dictator in RPV5-5 will transfer less money compared to the corresponding IDG5-5 and also less money in RPV1-9 than in the corresponding IDG1-9.

In the comparison between the two treatments, RPV1-9 and RPV5-5, many conflicting factors are at stake. As the initial endowment on the account of the recipient is made by the computer, emotions should play no significant role. On the other hand, there is empirical evidence that transfers are higher in *take* options like RPV1-9 than in *give* options, but it remains unclear if individuals perceive RPV5-5 as a *take* option, as a *give* option or as neither nor. The status quo bias also supports higher transfers in RPV1-9, while RPV5-5, with its initial equal distribution, supports the *equality* principle of justice and thus higher transfers. Following the further argument that there is no objective difference between RPV5-5 and RPV1-9, this leads to:

Hypothesis H2c: If the distribution is suggested by the computer in a way recipients have to reallocate money in some way, there is no significant difference in transfer rates between RPV5-5 and RPV1-9.

Method

Experimental design

To test our hypotheses, we keep the standard EDG design of Forsythe et al. (1994) regarding the total control of the dictator over the distribution of the endowment between her and the recipient and regarding anonymity in a laboratory setting. After the decision and after the participants saw their game's financial outcomes on the computer screens in the laboratory, participants were asked to rate the reasons for their decision on a six-level scale for the following items: *outcome maximization*, *fairness*, *second-order beliefs*, and *emotion*. To control the *need* principle of justice, dicta-

tors, as well as recipients, are given the option to transfer money to an external charity as well as to the other player.

Inverse Dictator Game (IDG) (H1): In an anonymous laboratory setting the recipient is asked on her computer screen in strategy mode (Selten, 1967; Mitzkewitz, & Nagel, 1993), which division (in 1 EUR increments) of the 10 EUR endowment she wants to propose to the dictator. The dictator is asked on her screen in strategy mode (Selten, 1967; Mitzkewitz, & Nagel, 1993) for each of the 11 possible propositions of the recipient, which division of the money she selects. For instance, if the recipient proposes a division of 6 EUR for the dictator and 4 EUR for herself and if the dictator chooses a division of 8 EUR for herself and 2 EUR for the recipient as a response to the proposition of the recipient, this specific answer of the dictator to the proposition of the recipient is realized. Following **H1**, we predict that the transfers of the dictator will be higher than in the standard EDG.

To gain more information about the decision behavior, we ask the recipients which division they would have proposed as a second choice, a third choice, and so forth. After conducting the experiment, we asked both the dictator and the recipient, which reasons led to their decisions. We hypothesize that fairness as a decision reason would be more important than self-interest.

For H2a, H2b, and H2c, in the RPV condition, the dictator gets the information, that the endowment of 10 EUR is initially distributed between the dictator's account and the account of the recipient as follows: 5-5 (condition RPV5-5), which means that 5 EUR are on the account of the dictator and 5 EUR on the account of the recipient; or 1-9 (condition RPV1-9), which means that the dictator has initially 1 EUR, and the recipient has 9 EUR on her account. Though, the dictator is free to transfer any amount of money to herself or the recipient by reallocating any possible division of the 10 EUR on her account and the recipient's account (in 1 EUR increments). Following **H2a**, we predict that in the RPV treatments the amount the dictator gives to her recipient is significantly higher than in the standard EDG benchmark. In RPV5-5, transfers shall be lower than in IDG5-5. The same applies to RPV1-9 and IDG1-9. Strictly speaking, we look at the following values in our analysis: the outcome of the RPV treatments RPV5-5 and RPV1-9, respectively, and the reaction of the dictator to the recipient's proposal in strategy mode in IDG (react), which is react5-5 and react1-9 (**H2b**). Furthermore, we predict no significant difference between RPV5-5 and RPV1-9 (**H2c**). The decision reason should be more self-interest than in the corresponding IDG condition.

Experimental procedures

We recruited a total of 72 participants (with additional six reserve players) from the participant pool of the Essen Laboratory for Experimental Economics (elfe) of the University of Duisburg-Essen by using the recruitment software ORSEE (Greiner, 2015). We ensured that the participants did not attend an EDG before and that we had

an equal gender quota. In a total of 12 sessions, all three design variants of the EDG were played simultaneously (IDG, RPV5-5, RPV1-9).

In each session, two participants played the IDG, two the RPV5-5, and two the RPV1-9 treatment. Thus, three participants were assigned the role of the dictator in their respective treatment. They were paired with three participants who were assigned the role of recipients. All participants read the written instructions (German version in the appendix). Thereafter, comprehension questions could be asked and the experimenter would see each participant individually in the soundproof cabins to answer the questions. Then the game appeared on the computer screen in each cabin. After the decision, the participants were asked to report the reasons for their decision. Following this, participants had to rate the importance of the following criteria for their decision on a six-level scale: *outcome maximization*, *fairness*, *second-order beliefs*, and *emotional arousal*. Then the participants learned their payoff on the computer screen. The experiment costed a total of 555 EUR, resulting from 36 dictator games multiplied by 10 EUR plus 78 times 2.50 EUR as the show-up fee. From a total of 555 EUR, only participants of the experiment were paid.

Results

Descriptive statistics

The experiment was conducted in 12 sessions with 6 participants each and an average duration of approximately 23 minutes. Among the 72 participants, 67 participants had not participated in any laboratory experiment before, while the other five experienced participants had not participated in any form of EDG. The gender distribution of male/female participants was in treatment IDG 10/14, in treatment RPV5-5 9/15, and in treatment RPV1-9 14/10. The participating undergraduate students study business/economics (27), MINT (23), teaching (8), social sciences (4), other areas of studies (10).

Table 1 Demographics of participants (Results for age, semester, and final profit are means with Standard deviation in parentheses). Final profit designates the sum each participant receives after the experiment, including show-up fee minus the amount she gave to the other players and the charity “Médecins Sans Frontières⁴” plus the amount she receives from the other participants. In the lines underneath the final profit is split between the group of dictators and the group of recipients.

Treatment	<i>n</i>	IDG		RPV5-5		RPV1-9	
		<i>M(SD)</i>	<i>n</i>	<i>M(SD)</i>	<i>n</i>	<i>M(SD)</i>	<i>M(SD)</i>
Age	24	21.83 (3.13)	24	21 (2.86)	24	22.5 (3.16)	
Semester	24	1.33 (.82)	24	1.58 (1.38)	24	1.96 (1.33)	
Final Profit in EUR	24	6.37 (2.05)	24	6.57 (2.32)	24	6.54 (2.48)	
Final profit Dictators in EUR	12	6.1 (1.96)	12	7.9 (2.08)	12	8 (1.9)	
Final profit Recipients in EUR	12	6.63 (2.19)	12	5.25 (1.75)	12	5.08 (2.15)	

In Table 1, the demographics of all participants divided by treatments are shown as well as the final profits. The treatment differences regarding final profit are clearly visible.⁵ The laboratory experiments were performed with zTree (Fischbacher, 2007). All session data was then analyzed with the help of the software SPSS and STATA. In order to compare our results with relevant dictator game experiments in the literature, we also reanalyzed raw data of Forsythe et al. (1994), Bardsley (2008), and Yamamori et al. (2008).

In Figure 2, we present our main results:

4 Donations were made to “Ärzte ohne Grenzen,” the German section of “Médecins Sans Frontières”.

5 After the dictator’s decision, all participants could give their money to the charity “Médecins Sans Frontières”. In total, the 72 participants gave 72.47 EUR.

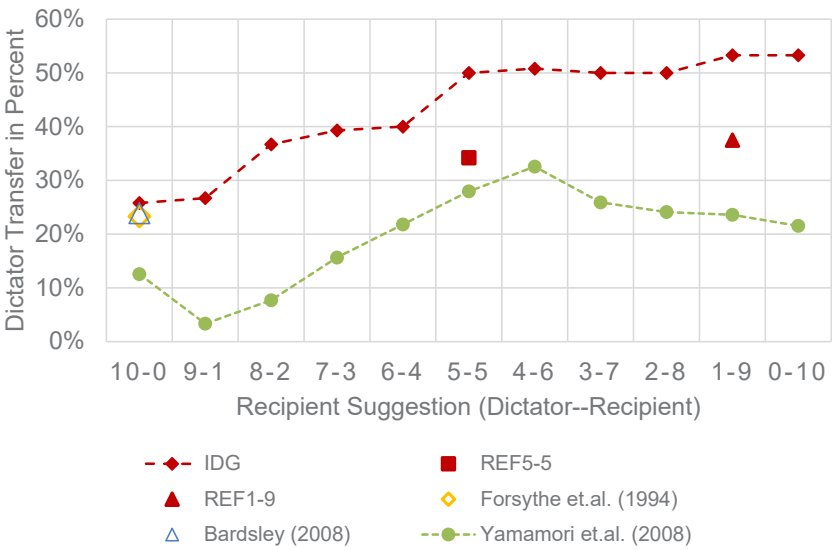


Figure 2 This figure shows the main results of our treatments as compared with the results of Forsythe et al. (1994), Bardsley (2008), and Yamamori et al. (2008). We show average dictator transfers given the suggestions by the recipient. Thereby, 10-0 represents the suggestion that 10 EUR will be given to the dictator, and 0 EUR will be given to the recipient and so forth.

As Figure 2 illustrates, in IDG strategy mode dictators decide to transfer 25.8 % to the recipient if she demands zero. This result is in line with the average transfer rates in the literature. Our 10-0 condition is equivalent to the standard dictator game with the framing that all funds are initially in possession of the dictator. For instance, in Forsythe et al. (1994) dictators transfer, on average, 23.3 %, though less than 28.35 % (Engel, 2011). Figure 2 illustrates that even though the objective situation does not change, the transfer from dictator to recipient increases with the recipient’s demand. Finding a climax with the condition 1-9 and 0-10, where the recipient suggests the distribution of 90 % respectively 100 % on her account. The average transfer in this case towards the recipient is 53.3 %, which is higher than the average transfer in the benchmark EDG of Forsythe et al. (1994) and even higher than the Ultimatum Game transfer in Forsythe et al. (1994) with 46.7 %.

In contrast to the findings of Yamamori et al. (2008), the transfers are higher the more impudent the recipient is. For instance, if the recipient suggests in the IDG 10 EUR for the dictator and 0 EUR for herself, the dictator follows this proposition and only gives an average of 2.58 EUR. However, the more the recipient demands, the more she gets. This result seems to suggest that there is no punishment for extreme demands by the recipient on an aggregate level in our neutral design which does neither use negative framing (“minimal acceptable offers”) nor an additional “not to tell” option as in Yamamori et al. (2008).

Table 2 shows the impact of the recipient’s hypothetical proposition (strategy mode) on the dictator’s decision for each of her 11 possible proposals:

Table 2 IDG treatment: Decision of Player A (Dictator) in strategy mode in response to Player B (Recipient)

Player B's suggestion	10-0	9-1	8-2	7-3	6-4	5-5	4-6	3-7	2-8	1-9	0-10
Average	2.58	2.67	3.67	3.92	4	5	5.08	5	5	5.33	5.33
Standard deviation	2.109	1.723	1.497	1.621	1.414	.000	.515	1.206	1.809	2.060	2.425
First Quartile	.25	1	2	3	4	5	5	5	5	5	5
Median	2	2.5	4.5	4.5	4	5	5	5	5	5	5
Third Quartile	5	4.75	5	5	5	5	5	5	5	5	5

In RPV5-5 and RPV1-9, we can show that the average transfer is higher than in the regular EDG⁶. In Table 3, we present the medians of player A's (dictator) decisions in both RPV-treatments:

Table 3 RPV treatments: Decision of Player A (Dictator)

Treatment	RPV5-5	RPV1-9
Average	3.42	3.75
Standard deviation	1.676	1.765
First Quartile	2	2.5
Median	3.5	4.5
Third Quartile	5	5

Individual decision patterns

In the IDG strategy mode, several dictator player types can be identified: 16.7 % of the players choose a consistent solution, here in the form of *constant equal split*. 41.7 % can roughly be assigned to the strategy *Follow suggestions, but only until the equal split*. 16.7 % give the recipient as much as she wants (up to 100 %). 25.0 % cannot be assigned to a consistent pattern (see the IDG individual decision patterns in strategy mode in appendix II). The central result is: No player chooses the game-theoretic, maximum payout solution (constant 0 EUR transfer) and no player punishes

6 In the RPV5-5 treatment versus the RPV1-9 treatment, it can be argued (according to Zhang & Ortmann, 2014, in contrast to Engel, 2011) that the dictators in RPV5-5 take less from the recipients (1.58) than in the treatment RPV1-9 (5.25).

impudent suggestions of the recipient. Despite the inconsistent decision patterns in strategy mode, the equal split of 5 EUR was realized for all 24 IDG players.

Mean Comparisons

In the IDG, the recipient suggests an allocation to the dictator first before the dictator makes the final allocation decision. Our results provide evidence in support of hypothesis **H1a**. For instance, the IDG achieves higher transfers compared to the benchmarks in the literature (Forsythe et al., 1994; Bardsley, 2008). Specifically, every single suggestion by the recipient was responded to with, on average, higher transfer decisions by the dictators in comparison to the EDG baselines.

As an example, when the recipient suggests 0-10, that is the recipient suggests that 0 EUR are transferred to the dictator and 10 EUR are transferred to herself, then the mean response by the dictators is 5.33 EUR with a 10 % confidence interval of [4.08;6.59] which excludes the EDG baselines. That is, we accept the alternative hypothesis that our average transfers are higher than the benchmarks in the literature for the suggestion 0-10. The same is true for all other suggestions by the recipient except for 10-0 and for 9-1. That means, when the recipient asks very little for herself (0 EUR and 1 EUR), then the average transfers by the dictators were statistically not different from the EDG baselines. Although even then, the average transfers were higher than the EDG baseline.

We see further evidence that the dictator's average transfer rates increase according to the suggested transfer by the recipient. Specifically, in Figure 3, the dictators' transfer means monotonically increase with the amounts the recipients request for themselves. These results can be described with the well-known proverb "The squeaky wheel gets the grease." The more the recipients demand, the more they are given by the dictators.

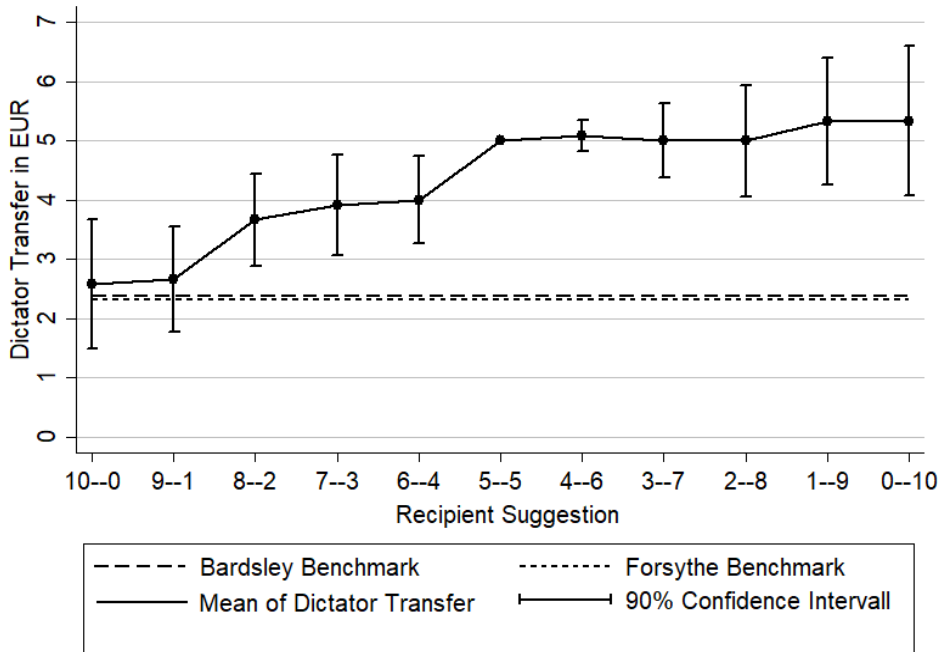


Figure 3 This Figure shows the means of the dictator's transfers in relation to the suggestions by the recipient together with 90 % confidence intervals using critical t-values. The figure also shows the baseline by Forsythe et al. (1994) and by Bardsley (2008).

Next, we pool together all those suggestions by the recipient proposing less than 50 % of the allocation to themselves. Specifically, these suggestions are 10-0, 9-1, 8-2, 7-3, and 6-4. We also pool together all suggestions by the recipient proposing more than 50 % of the allocation to themselves. Specifically, these suggestions are 0-10, 1-9, 2-8, 3-7, and 4-6. Figure 4 represents the result graphically. When the recipients suggest less than 50 % for themselves, the mean allocation from the dictators is 3.64. When the recipients suggest more than 50 % for themselves, the mean allocation from the dictators increases to 5.15, which is statistically significantly higher than 3.64 as indicated by the 90 % confidence intervals.

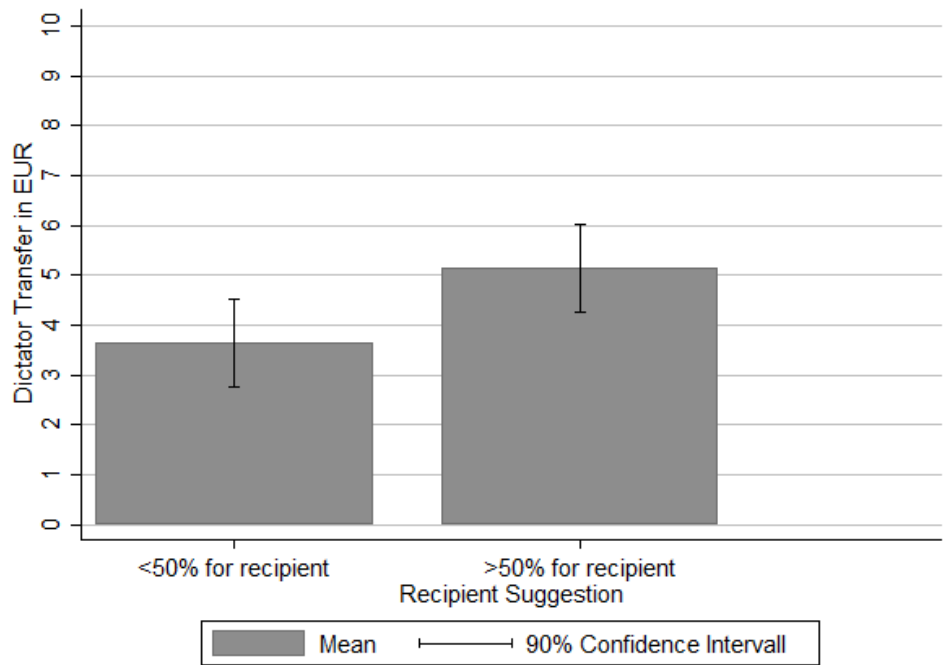


Figure 4 This Figure shows the means of the dictator’s transfers in response to the recipient’s suggestions together with 90 % confidence intervals using critical t-values. We pool together the recipients’ suggestions proposing less than 50 % of the allocation for themselves (<50 % for the recipient) and those proposing more than 50 % of the allocation for themselves (>50 % for the recipient).

In addition, we apply the non-parametric Mann-Whitney U test in order to determine whether or not the dictators’ responses to various suggestions by the recipients differ systematically. Specifically, we test whether the dictator responds to various suggestions by the recipient with the same distribution. Our main findings are reported in Table 4. Specifically, Table 4 reports the p-values of dictator transfers given all permutations of pairwise comparisons between the suggestions by the recipient. For example, the distribution of dictator transfers if the recipient suggests 19 is statistically highly different from the distribution when the recipient suggests 91 (p-value = 0.002).

We color the table in the style of a “heat map.” That means, the smaller the p-value (the stronger the statistical significance), the darker the blue tone of the cell of the p-value. What we can show in this way is that the further the recipients’ suggestions are apart from each other, the more likely it is that the Mann-Whitney U test rejects the null hypothesis that the two samples come from the same distribution. In other words, Table 4 delivers further empirical evidence that the suggestions by the recipient strongly influence the allocation decisions by the dictators.

For example, the dictators’ allocation decisions are not statistically different from one another, whether the recipients suggest 9-1 or 8-2 with a p-value of 0.105. That

means, whether the recipient suggests 1 EUR or 2 EUR for herself does not make a big difference to the dictator's transfer. Comparing the suggestions 9-1 to 7-3, 6-4 delivers a statistically significant p-value at the 10 % significance level. That means, whether the recipient suggests 1 EUR for herself or 3 or 4 EUR seems to influence the decisions by the dictator. The distribution of transfer decisions is highly significant at the 1 % level. All results and permutations are reported in Table 4.

Table 4 This Table reports the p-values based on the Mann-Whitney U test of dictator transfer decisions given all permutations of pairwise comparisons between the suggestions by the recipient. The smaller the p-value, the darker the blue tone.

Mann-Whitney U Test p-values		Recipient Suggestion (Dictator-Recipient)										
		10-0	9-1	8-2	7-3	6-4	5-5	4-6	3-7	2-8	1-9	0-10
Recipient Suggestion (Dictator-Recipient)	10-0		0.883	0.159	0.107	0.150	0.001	0.002	0.004	0.008	0.004	0.006
	9-1			0.105	0.084	0.074	0.000	0.001	0.002	0.004	0.002	0.004
	8-2				0.521	0.735	0.006	0.010	0.028	0.058	0.028	0.028
	7-3					0.927	0.031	0.033	0.099	0.165	0.057	0.061
	6-4						0.002	0.006	0.050	0.058	0.013	0.014
	5-5							0.547	1.000	1.000	0.547	0.547
	4-6								0.774	0.774	0.909	0.909
	3-7									1.000	0.666	0.666
	2-8										0.666	0.666
	1-9											0.970
	0-10											

Regression Analysis

Table 5 shows the results of our regression analysis using ordinary least squares (OLS). The dependent variable is the dictator's decision about how much money to transfer to the recipient. We regress the dictators' decisions on our key independent variable, which is the suggestion by the recipient. We also include typical controls as independent variables such as the self-reported age of the dictators, the gender, as well as the self-reported reasons for their decision-making, including *outcome maximization*, *fairness*, *second-order beliefs*, and *emotions*.

Table 5 shows three specifications of our empirical model (I) to (III). In total, we have 132 observations, and the model fits reasonably well to the data with an adjusted coefficient of determination of 0.33 in specification (III).

In all specifications (I) to (III), the recipients' suggestion is statistically significant at the 1 % confidence level with a coefficient of 0.288. That means, on average, if the recipient asks for 1 EUR more (holding all other variables constant), the dictator increases her transfer by 0.29 EUR.

We observe statistically significant age and gender effects in specification (II) and (III). In terms of age, we find that the older the dictators, the more they transfer to the recipients. This result is highly significant at the 1 % level in both specifications. Specifically, the coefficient of 0.126 in specification (III) tells us that if a dictator

is one year older than another dictator (all else equal), the older dictator transfers 0.13 EUR more.

Table 5 This Table reports the output of our regression analysis. The independent variable in all three models is the dictator transfers, and the main independent variable is the suggestion by the recipients. In parentheses are the values of the t-statistics. We denote * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Regression Table			
	(I)	(II)	(III)
Dependent variable: Dictator's transfer			
Recipient's suggestion	0.288*** (6.47)	0.288*** (6.88)	0.288*** (6.88)
Age		0.143*** (3.57)	0.126*** (2.59)
Gender (0=male;1=female)		1.284*** (3.94)	0.988*** (2.68)
out_max			0.121 (0.96)
Fairness			-0.425 (-1.09)
snd_order			0.085 (0.72)
Emotion			0.169 (1.11)
_cons	2.886*** (10.97)	-1.155 (-1.10)	0.378 (0.17)
Observations	132	132	132
R ²	0.24	0.34	0.36
Adj. R ²	0.24	0.33	0.33

In terms of gender, we find that female dictators transfer more on average than male dictators. This result is highly significant at the 1 % level in both specifications where we control for gender. Specifically, the coefficient of 0.988 in specification (III) tells us that female dictators, on average, transfer almost 1 EUR more than male dictators.

We find no statistically significant effects by the variables *outcome maximization*, *fairness*, *second-order beliefs*, and *emotions*.

The regression outputs deliver further evidence that the suggestions by the recipients significantly influence the decision of dictators. The more the recipients request, the more the dictators transfer.

Hypotheses 2

Dictators transfer an average of 34.2 % of their endowment in RPV5-5 (with initially 5 EUR in the recipient's account) and 37.5 % in RPV1-9 (with initially 9 EUR in the recipient's account) to the recipient. Both transfer rates are higher compared to an average of 23.3 % transfers in the benchmark EDG of Forsythe et al. (1994). With a Mann-Whitney U test, we find significant differences between Forsythe et al. (1994) and RPV1-9 with $U = 82.000$ and $p = .038$. For Forsythe et al. (1994) and RPV5-5, we find no significant differences.

Consequently, **H2a** holds true for RPV1-9 with significantly higher transfers but not for RPV5-5. Regarding **H2c**, we do not find significant differences between the treatments RPV1-9 and RPV5-5 regarding the dictator decision but also regarding decision reasons like *fairness*, *outcome maximization*, *emotions*, or *second-order beliefs*. To test differences between the RPV treatment and the IDG treatment, we analyze the outcome of the RPV treatments, respectively, and the reaction of the dictator to the recipient's proposal in strategy mode in IDG (react), that is react5-5 and react1-9 (**H2b**).

We find significant differences between RPV1-9 and react1-9 ($p = .045$). Respectively, we find significant differences between RPV5-5 and react5-5 ($p = .014$). These differences result from the fact that the dictators in the IDG react5-5 (strategy mode) follow the recipient's proposal of equal distribution in 100 % of the cases. In contrast, in the RPV5-5, only 41.7 % of the dictators choose the equal distribution. The remaining dictators prefer to give less than the equal distribution. In the IDG condition react1-9 (strategy mode), two dictators (16.7 %) follow the proposition of the recipients suggesting 9 EUR for the recipient and 1 EUR for the dictator. 75 % of the dictators choose the equal distribution and one (8.3 %) the distribution 9-1. In the RPV1-9 condition, none of the dictators follows the distribution proposition 1-9. Half of the dictators choose the equal distribution, the other half give less than equal distribution to the recipient.

Discussion

Several of our empirical findings are in line with the previous literature on similar experimental settings. Many of our results support our predictions, which we derived from the model of frame selection (MFS) and from the assumption that strong reference points and social situations both lead to higher transfer rates. To the best of our knowledge, the paper at hand is the first attempt to explain findings in an experimental dictator game with the MFS. These results will be discussed in the following

section. Thereafter, we discuss the highly significant finding that if recipients ask for more for themselves than what would be left for the dictator, this leads to increasing transfer rates above the equal split, both, on an individual and on an aggregate level, contrary to the findings of Yamamori et al. (2008). The differences in the empirical findings may be the result of cultural differences. The adage “The squeaky wheel gets the grease”, with which we described our results, reflects the US-American and also the German⁷ culture, whereas the Japanese culture knows the proverb “The nail that stands out gets pounded down”, which means the opposite. As predicted from an MFS perspective, in the variant of the *Reference Point Variation* (RPV) with two different *take* options given by the computer, dictators transfer an average of 34.2 % of their endowment to the recipient in RPV5-5 and 37.5 % in RPV1-9. Both transfer rates are significantly higher compared to an average of 23.3 % transfers in the benchmark EDG of Forsythe et al. (1994). In the variant of the EDG, the *Inverse Dictator Game* (IDG), a social situation with more right[s] to choose and thereby corresponding principles of justice being induced, the dictators tend to follow the suggestions made by the recipient, as predicted. Mean and pooled mean comparisons confirm these results from descriptive statistics. The OLS regression analysis with 132 observations shows in all model specifications (I) to (III) that the recipient’s suggestion is statistically significant at the 1 % confidence level with a coefficient of 0.288. The findings of higher transfer rates in RPV and transfer rates in IDG following suggestions of recipients up to the equal split are in line with results in similar experimental designs (voice, rudimentary communication, e.g., Yamamori et al., 2008). Sauermann (2017) also demonstrates the importance of distributive fairness in a majority voting experiment. Nevertheless, contrary to these similar designs in the literature, the more a recipient asks for, the more she gets, even beyond the equal split. We analyze these descriptive results on an aggregate and also on an individual level and show that they are robust for all statistical procedures we performed. This leads to the conclusion that in our design where a social situation is strongly induced, recipients do not punish impudent suggestions and partly follow even greedy suggestions: “The squeaky wheel gets the grease”.

In strong social relations involving fear (e.g., military) or trust (e.g., family), following others’ suggestions can be theoretically explained by emotional or habit-driven behavioral scripts. However, in an anonymous laboratory setting it remains unclear why these mechanisms should be at stage to crowding out principles such as equality or reciprocity.

The results lead to one of the core discussions in experimental economics why tiny variations in design can lead to major changes in behavioral patterns on an individual level and thus to extreme variations in aggregated experimental outcomes. So far, one approach is to collect more and more experimental data in order to see the bigger picture and develop matching theoretical models. Thus, we suggest involving the core elements of the MFS, namely the *match* in future research, because the

7 Das Rad, das am lautesten quietscht, bekommt das meiste Fett.

match has already proven that it is capable of providing a theoretical background for the heterogeneous findings.

Literature

- Abbink, K., & Sadrieh, A. (2009). The pleasure of being nasty. *Economics Letters*, 105(3), 306-308. doi:10.1016/j.econlet.2009.08.024
- Bardsley, N. (2008). Dictator game giving: Altruism or artefact?. *Experimental Economics*, 11(2), 122-133. doi: 10.1007/s10683-007-9172-2
- Bergh, A. & Wichardt, P. C. (2018). *Mine, ours and yours? Unintended framing effects in dictator games* (CESifo Working Paper No. 7049). Munich: Munich Society for the Promotion of Economic Research - CESifo GmbH.
- Biniossek, C., & Betz, D. (2018). *Der Einfluss unterschiedlicher Gerechtigkeitsprinzipien auf Mehrheitsentscheidungen in einem spieltheoretischen Social-Good-Experiment*. (GESIS Papers, 2018/20). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. doi: 10.21241/ssoar.58261
- Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1), 166-193. doi: 10.1257/aer.90.1.166
- Brosig-Koch, J., Riechmann, T., & Weimann, J. (2007). *Selfish in the end? An investigation of consistency and stability of individual behavior* (FEMM Working Paper No. 05). Magdeburg: Otto-von-Guericke-University Magdeburg.
- Cartwright, E. (2016). A comment on framing effects in linear public good games. *Journal of the Economic Science Association*, 2(1), 73-84. doi:10.1007/s40881-016-0024-8
- Cox, J. C., Friedman, D., & Sadiraj, V. (2008). Revealed altruism. *Econometrica*, 76(1), 31-69. doi:10.1111/j.0012-9682.2008.00817.x
- Cook, K. S., & Hegtvedt, K. A. (1983). Distributive justice, equity, and equality. *Annual review of sociology*, 9(1), 217-241.
- Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice?. *Journal of Social issues*, 31(3), 137-149.
- Dreber, A., Ellingsen, T., Johannesson, M., & Rand, D. G. (2013). Do people care about social context? Framing effects in dictator games. *Experimental Economics*, 16(3), 349-371. doi: 10.1007/s10683-012-9341-9
- Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and economic behavior*, 47(2), 268-298. doi:10.1016/j.geb.2003.06.003
- Edwards, W. (1954). The theory of decision making. *Psychological bulletin*, 51(4), 380-417. doi: 10.1037/h0053870
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4), 583-610. doi: 10.1007/s10683-011-9283-7
- Esser, H. (1996). Die Definition der Situation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 48(1), 1-34.

- Esser, H. (1999). *Soziologie. Allgemeine Grundlagen*. Frankfurt/New York: Campus Verlag.
- Esser, H. (2001a). *Soziologie. Spezielle Grundlagen* (Vol. 1). Frankfurt a.M./New York, NY: Campus.
- Esser, H. (2001b). *Soziologie. Spezielle Grundlagen* (Vol. 6). Frankfurt a.M./New York, NY: Campus.
- Esser, H. (2002a). In guten wie in schlechten Tagen? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 54, 27–63.
- Esser, H. (2002b). Ehekrisen: Das (Re-) Framing der Ehe und der Anstieg der Scheidungsraten. *Zeitschrift für Soziologie*, 31(6), 472–496. doi: 10.1515/zfsoz-2002-0602
- Esser, H. (2003). Der Sinn der Modelle. Antwort auf Götz Rohwer. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 55(2), 359–368. doi:10.1007/s11577-003-0055-z
- Esser, H. (2010a). Das Modell der Frame-Selektion. Eine allgemeine Handlungstheorie für die Sozialwissenschaften? In G. Albert & S. Sigmund (Eds.), *Soziologische Theorie kontrovers. Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie* (pp. 45–62).
- Esser, H. (2010b). Sinn, Kultur, Verstehen und das Modell der soziologischen Erklärung. In M. Wohlrab-Sahr (Ed.), *Kultursoziologie* (pp. 309–335). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/978-3-531-92300-0_13
- Esser, H. (2017). *Rational Choice or Framing? RCT and MFS as Two Approaches to Explain the Patterns in the Fehr-Gächter-Experiments on Cooperation and Punishment in the Contribution to Public Goods*. Unpublished manuscript.
- Esser, H. (2018). Sanktionen, Reziprozität und die symbolische Konstruktion einer Kooperations-„Gemeinschaft“. *Zeitschrift für Soziologie*, 47(1), 8–28. doi: 10.1515/zfsoz-2018-1001
- Esser, H. (2019) Rational Choice oder Framing? In J. Mayerl, T. Krause, A. Wahl, & M. Wuketich (Eds.), *Einstellungen und Verhalten in der empirischen Sozialforschung* (pp. 35–64). doi:10.1007/978-3-658-16348-8_3
- Esser, H., & Kroneberg, C. (2015). An integrative theory of action. The model of frame selection. In E. J. Lawler, S. R. Thye, & J. Yoon (Eds.) *Order on the Edge of Chaos. Social Psychology and the Problem of Social Order* (pp. 63–85). New York, NY: Cambridge University Press.
- Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2), 293–315. doi:10.1016/j.geb.2005.03.001
- Fehr, E., & Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817–868. doi: 10.1162/003355399556151
- Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 10(2), 171–178. doi: 10.1007/s10683-006-9159-4
- Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Kiesel, A., Kunde, W., Mayr, S. Moeller, B., Möller, M, Pfister, R., & Philipp, A. (2020). Binding and Retrieval in Action Control (BRAC). *Trends in Cognitive Sciences*, 1–13. doi: 10.1016/j.tics.2020.02.004.
- Frith, C. D., & Haggard, P. (2018). Volition and the brain – revisiting a classic experimental study. *Trends in Neurosciences*, 41(7), 405–407. doi: 10.1016/j.tins.2018.04.009

- Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3), 347-369. doi:10.1006/game.1994.1021
- Gerlach, P., & Jaeger, B. (2016). Another frame, another game? Explaining framing effects in economic games. In A. Hopfensitz, & E. Lori (Eds.), *Proceedings of norms, actions, games (NAG 2016)*. Toulouse: Institute for Advanced Studies. doi: 10.17605/OSF.IO/AB5YP
- Goerg, S. J., Rand, D., & Walkowitz, G. (2019). Framing effects in the prisoner's dilemma but not in the dictator game. *Journal of the Economic Science Association*, 1-12. doi: 10.1007/s40881-019-00081-1
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-125. doi: 10.1007/s40881-015-0004-4.
- Grossman, P. J., & Eckel, C. C. (2015). Giving versus taking for a cause. *Economics Letters*, 132, 28-30. doi:10.1016/j.econlet.2015.04.002
- Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental brain research*, 126(1), 128-133.
- Heinrich, T., Riechmann, T., & Weimann, J. (2009). *Game or frame? Incentives in modified dictator games* (FEMM Working Paper No. 09008). Magdeburg: Otto-von-Guericke-University Magdeburg.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *The journal of business*, 59(4), 285-300.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1), 193-206. doi: 10.1257/jep.5.1.193
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291. doi:10.2307/1914185
- Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. In A. Kurtz (Ed.), *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere* (284(1), pp. 1-17). Springer. doi: 10.1007/BF00412364
- Kroneberg, C. (2005). Die Definition der Situation und die variable Rationalität der Akteure. Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie*, 34(5), 344-363. doi: 10.1515/zfsoz-2005-0502
- Kroneberg, C. (2011). *Die Erklärung sozialen Handelns: Grundlagen und Anwendung einer integrativen Theorie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kroneberg, C. (2014). Frames, scripts, and variable rationality: An integrative theory of action. In G. Manzo (Ed.), *Analytical Sociology. Actions and Networks* (pp. 97-123). Chichester, UK: Wiley.
- Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3), 495-524. doi: 10.1111/jeea.12006
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The uncon-

- scious initiation of a freely voluntary act. *Brain*, 106(3), 623-642. doi: 10.1093/brain/106.3.623
- List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political economy*, 115(3), 482-493. doi:10.1086/519249
- Mitzkewitz, M., & Nagel, R. (1993). Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory*, 22(2), 171-198.
- Nann, M., Cohen, L. G., Deecke, L., & Soekadar, S. R. (2019). To jump or not to jump - The Bereitschaftspotential required to jump into 192-meter abyss. *Scientific reports*, 9(1), 1-9. doi: 10.1038/s41598-018-38447-w
- Rawls, J. (1971). *A theory of justice*. Harvard university press.
- Rohwer, G. (2003). Modelle ohne Akteure. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 55(2), 340-358. doi: 10.1007/s11577-003-0054-0
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1), 7-59.
- Sauermann, J. 2017. Do Individuals Value Distributional Fairness? How Inequality Affects Majority Decisions. *Political Behavior*. doi: 10.1007/s11109-017-9424-6
- Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopol-experiments. In H. Sauermann (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung* (pp. 136-168).
- Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2), 274-279.
- Smith, V. L. (2010). Theory and experiment: What are the questions? *Journal of Economic Behavior*, 73(1), 3-15. doi: 10.1016/j.jebo.2009.02.008
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297-323. doi: 10.1007/BF00122574
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.
- Walkowitz, G. (2018). *On the validity of probabilistic (and cost-saving) incentives in dictator games: A systematic test* (MPRA Paper No. 89797). Munich: Munich Personal RePEc Archive. <https://mpra.ub.uni-muenchen.de/89797/>
- Yamamori, T., Kato, K., Kawagoe, T., & Matsui, A. (2008). Voice matters in a dictator game. *Experimental Economics*, 11(4), 336-343. doi: 10.1007/s10683-007-9168-y
- Zhang, L., & Ortmann, A. (2014). The effects of the take-option in dictator-game experiments: A comment on Engel's (2011) meta-study. *Experimental Economics*, 17(3), 414-420. doi: 10.1007/s10683-013-9375-7

Appendix I

Instructions (German):

Willkommen zum Experiment!

Vorbemerkung

Sie nehmen an einer Untersuchung des Entscheidungsverhaltens teil. Wir bitten Sie, nicht zu essen, nicht zu trinken, Ihr Handy auszuschalten und nicht mit anderen Teilnehmern zu sprechen. Während der Untersuchung werden Sie und die anderen Teilnehmer gebeten, Entscheidungen zu treffen. Dabei können Sie Geld verdienen. Wieviel Geld das sein wird hängt von Ihren Entscheidungen und den Entscheidungen der anderen Teilnehmer ab. Am Ende des Experimentes wird Ihnen Ihr gesamter Verdienst bar ausgezahlt. Hinzu kommt eine Teilnehmerpauschale in Höhe von 2,50 EURO. Bleiben Sie daher bitte nach dem Experiment solange an Ihrem Platz sitzen, bis Ihre Platznummer aufgerufen wird.

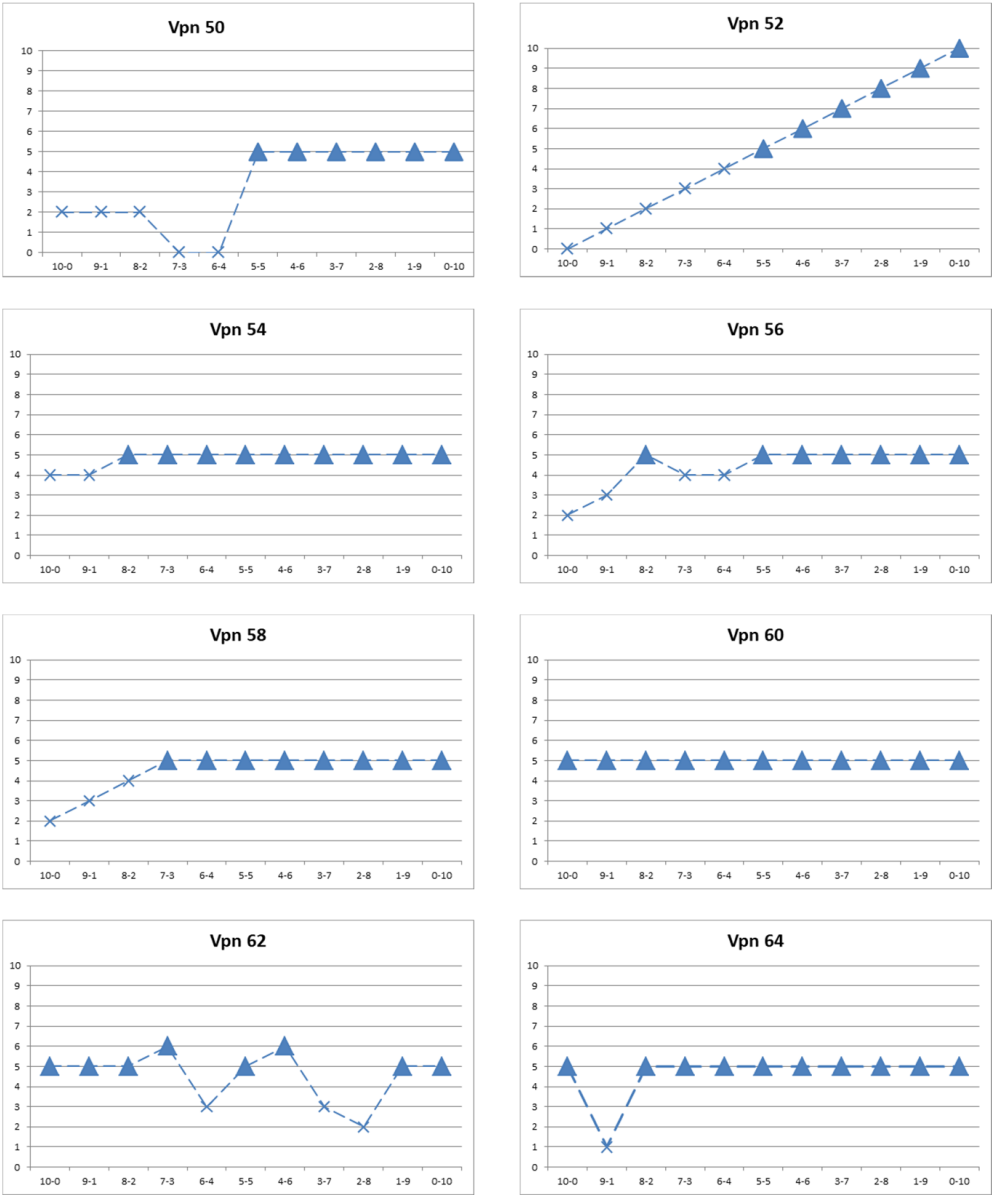
Entscheidungssituation

Das Experiment dauert etwa 30 Minuten. Vor jeder Entscheidung erhalten Sie detaillierte Instruktionen. An jeder Entscheidungssituation ist genau ein anderer Teilnehmer beteiligt. **Keiner der Teilnehmer wird im Verlauf des Experimentes Informationen über die Identität seines zugeteilten Partners oder die Identität der anderen Teilnehmer erhalten. Alle Entscheidungen werden also völlig anonym getroffen.**

Sollten Sie während des Experimentes Fragen haben, heben Sie bitte Ihre Hand, ein Mitarbeiter des Labors wird dann zu Ihnen kommen. **Jegliche Kommunikation mit den anderen Teilnehmern ist während des Experimentes untersagt.**

Appendix II

The following Figure presents the IDG individual decision patterns in strategy mode (y-axis: sum to recipient) depending on the recipients' suggestion (x-axis), whereas 10-0 means 10 EUR on the account of the dictator and 0 EUR on the account of the recipient. The figures show ▲ for values on the equal split (distribution of 5 EUR for the dictator and 5 EUR for the recipient) and above, while x for values underneath the equal split.



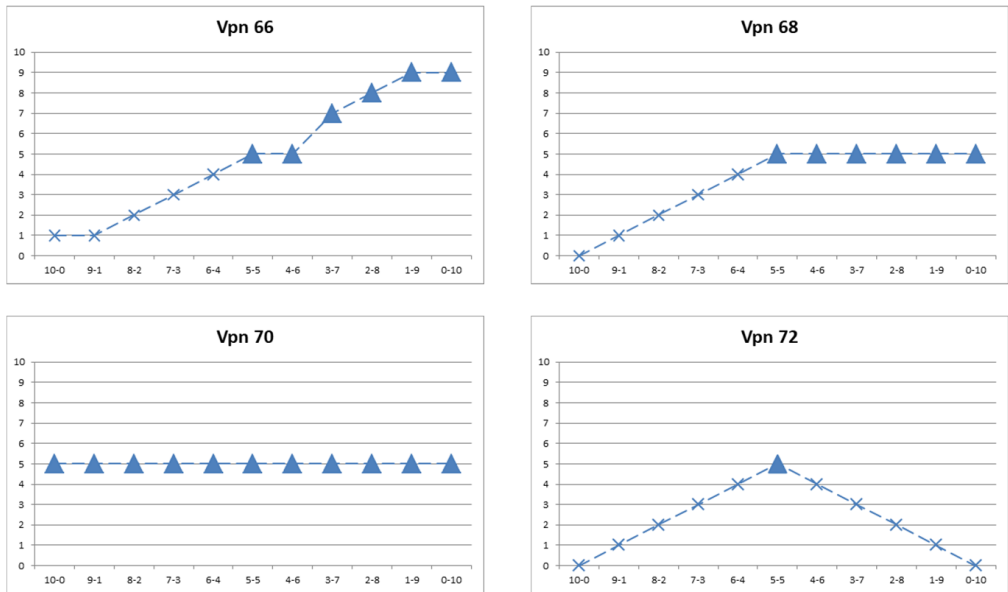


Figure 5 This Figures show the individual IDG dictators' decisions in strategy mode.

Die vorliegende Arbeit untersucht ausgewählte Theorien der Politikwissenschaft, Ökonomie, Soziologie und Psychologie in unterschiedlichen Anwendungsfeldern mit methodischen Instrumenten der experimentellen Spieltheorie und der empirischen Sozialforschung. Dabei ist die leitende Frage, ob die der Rational-Choice-Theorie (RCT) zugrunde liegende Annahme der Selektion der nutzenmaximalen Handlungsoption durch eine zentrale Exekutive eine ausreichend große Erklärungskraft für menschliches Entscheidungsverhalten hat. In den fünf Studien des Buches zeigt sich teils massive empirische Evidenz, die nicht im Einklang mit der RCT steht. Die Abweichungen von den durch die RCT prognostizierten Ergebnissen sind hierbei nicht zufällig, sondern systematisch. Die Befunde widersprechen den Prognosen der RCT umso mehr, je umfassender und kleingranularer Individuen im sozialen Feld betrachtet werden, je uneindeutiger und vom ökonomischen Kontext ferner eine Situation ist, je geringer die individuellen (finanziellen) Kosten sind (Low-Cost-Hypothese) und je stärker heuristische Informationsverarbeitungsprozesse beobachtbar sind. Daneben lassen sich in den Studien aber immer wieder auch eigen-nutzmaximierende Motive und rationales Handeln erkennen. Insbesondere im ökonomischen Kontext und auf Aggregatebene zeigt sich die große Leistungsfähigkeit der RCT und des ökonomischen Verhaltensparadigmas. Vor diesem Hintergrund wird gezeigt, dass Rahmentheorien und alternative Selektionsmechanismen diese zunächst so gegensätzlich scheinenden Befunde in einer neuen, höheren Einheit aufzuheben vermögen. In der Gesamtschau ermöglicht die methodische und theoretische Breite der Arbeit die Schlussfolgerung, die RCT als Spezialfall von Rahmentheorien zu konzeptualisieren und pragmatisch sowie anwendungsfall-spezifisch weiterhin zu nutzen.

This book examines selected theories from political science, economics, sociology, and psychology in different fields of applications using methodological tools from experimental game theory and empirical social research. The leading research question is whether the assumption of the underlying rational choice theory (RCT), for instance the selection of the utility-maximizing option by a central executive, has sufficient explanatory power for human decision behavior. Some of the five studies in this book reveal significant empirical evidence that is inconsistent with the RCT. Here, the deviations from the results predicted by the RCT are not random but systematic. The findings contradict the predictions of RCT more significantly, the more comprehensive and detailed individuals in the social field are observed, the more ambiguous and distant the situation from the economic context is, the lower the individual (financial) costs are (low-cost hypothesis), and the more heuristic information processes are observable. Furthermore, self-interest-maximizing motives and rational action can be identified recurrently. In the economic context and at the aggregate level, the great power of RCT and the economic behavioral paradigm becomes apparent. Against this background, it is shown that framing theories and alternative selection mechanisms can resolve these initially seemingly contradictory findings in a new, higher unity. Overall, the methodological and theoretical range of this research work allows the conclusion to conceptualize RCT as a special case of framing theories and to continue to use it pragmatically as well as context-specific.